

# **TBOTDETECTOR**

## **ANÁLISIS DE BOTS Y REDES DE BOTS EN LA RED SOCIAL TWITTER**

**Sergio Rodríguez Gundín**

Grado en Ingeniería del Software

Facultad de Informática

Universidad Complutense de Madrid



# **UNIVERSIDAD COMPLUTENSE MADRID**

TRABAJO DE FIN DE GRADO

2016 / 2017

Director: Jorge J. Gómez Sanz



## RESUMEN

Se presenta la herramienta Web para la identificación de perfiles bot *TBotDetector*, desarrollado para cualquier tipo de navegador web, adaptado para Smartphone y Tablet, se centra en la red social Twitter.

Twitter es una fuente de información de creciente importancia. Para el usuario es sencillo publicar contenido que sea visionado por el resto sin ser conscientes de la calidad, validez o su relevancia. Muchos usuarios son en realidad bots controlados por personas con el objetivo de promover o desacreditar contenidos. Cada vez es más difícil distinguirlos por lo que se hace necesario formas de luchar contra estas maniobras de contra información identificando y denunciando redes de bots.

Para ello, este proyecto plantea dos objetivos: decidir si una cuenta es gestionada o no por un bot y buscar conjuntos de bots que operen de forma conjunta. *TBotDetector* proporciona análisis del perfil y contenido de un set de tweets para detectar este tipo de usuarios.

El trabajo muestra cómo generar informes con evidencias que sugieran que una cuenta puede o no ser un bot y que puede o no pertenecer a una red bot. La fiabilidad de los informes no es absoluta, y por ello deben ser tomados como un punto de partida para otras acciones.

**Palabras clave** Aplicación Web, bot, redes bot, MeaningCloud, navegador Web, redes sociales, Twitter.

# ABSTRACT

TBotDetector is a web tool to identify bots in Twitter.

Twitter actually is an important source of information. Users find very easy to post their content so that other users can see them but they don't know the quality, validity or relevance of the information. Many users are not such, but programmed bots controlled by a human operator to promote or discredit content. It is increasingly difficult to distinguish them. For achieve this it is necessary to invent methods to identify and report them.

This Project has two objectives: the first is identify bots accounts and the second is discover botnets in Twitter. TbotDetector provides profile analysis and content a set of tweets.

The work shows analysis reports examples that indicate the possibility a user account is a bot or part of a net of bots. Reports are indicators, not facts. Therefore, they should be used only as a warning sign and the first step towards a more thorough investigation.

**Keywords** Web application, bot, botnet, MeaningCloud, browser, social networks, Twitter.



# Índice General

Capítulo 1 .....	1
INTRODUCCIÓN .....	1
1.1    Objetivos .....	2
1.2    Organización del documento .....	3
Capítulo 2 .....	4
ESTADO DEL ARTE .....	4
2.1    La red social Twitter .....	4
2.2    Aplicaciones para la detección de bots .....	6
2.3    Conclusiones.....	8
Capítulo 3 .....	9
REQUISITOS .....	9
3.1    Requisitos funcionales.....	10
3.1.1    Análisis Simple.....	11
3.1.2    Análisis múltiple .....	17
3.2    Requisitos no funcionales .....	24
Capítulo 4 .....	25
DISEÑO .....	25
4.1    Interfaz Gráfica.....	25
4.1.1    MeaningCloud .....	29
4.1.2    Arquitectura .....	34
4.1.3    Diagramas de secuencia.....	42
Capítulo 5 .....	52
EXPERIMENTOS .....	52
5.1    Análisis Simple.....	52
5.1.1    Experimentos con usuarios bot .....	52
5.1.2    Experimentos con usuarios no bot.....	61
5.2    Análisis Múltiple .....	69
5.2.1    Experimentos con redes bot .....	69
5.2.2    Experimentos sin redes bot.....	75
Capítulo 6 .....	77
6.1    Conclusión .....	77

6.2	Lineas futuras .....	77
6.3	Conclusion .....	78
	Bibliografía .....	79

# Índice de figuras

Figura 1. Relaciones posibles entre usuarios de Twitter .....	5
Figura 2. Casos de uso de la aplicación.....	10
Figura 3. Algoritmo principal para la detección de bots en el análisis simple.....	12
Figura 4. Extracción de clústeres más relevantes del usuario en función de la longitud o URL como similitud .....	13
Figura 5. Toma de decisión en función de los datos de la cuenta del usuario .....	14
Figura 6. Toma de decisión en función de determinadas características del usuario .....	15
Figura 7. Análisis de las categorías más destacadas del usuario mediante comparación de proporciones .....	16
Figura 8. Análisis de entidades más representativas y número de apariciones .....	16
Figura 9. Análisis de sentimiento y toma de decisión para el análisis en función de la polaridad obtenida .....	17
Figura 10. Algoritmo principal para la obtención de redes bot.....	18
Figura 11. Análisis de tweets duplicados y comparación entre el usuario bot y el analizado .....	19
Figura 12. Comprobaciones de la información de los clústeres para tomar una decisión .....	20
Figura 13. Comparación de entidades para el análisis de redes bot .....	21
Figura 14. Comparación de categorías para el análisis de redes bot .....	22
Figura 15. Análisis de la actividad de los usuarios y toma de decisión en base a los resultados obtenidos en la comparación .....	23
Figura 16. Vista inicial del análisis simple .....	26
Figura 17. Vista del menú desplegable desde Smartphone .....	26
Figura 18. Vista del análisis simple desde Smartphone.....	26
Figura 19. Vista del análisis simple visto en formato Tablet .....	27
Figura 20. Vista principal del análisis múltiple .....	28
Figura 21. Vista principal de la aplicación .....	29
Figura 22. Características extraíbles de las tecnologías MeaningCloud .....	31
Figura 23. Sistema de clasificación MeaningCloud.....	32
Figura 24. Arquitectura de la aplicación TBotDetector .....	35
Figura 25. Proceso de petición y respuesta gestionado por Spring MVC.....	36
Figura 26. Diagrama de clases encargadas de la configuración de la capa web .....	37
Figura 27. Diagrama de clases para el Análisis Simple .....	39
Figura 28. Diagrama de clases para el Análisis Múltiple.....	41
Figura 29. Diagrama de secuencia general para el Análisis Simple .....	42
Figura 30. Diagrama de secuencia Recuperar Información de Cuenta .....	43



Figura 31. Diagrama de secuencia Procesar Información de Cuenta .....	44
Figura 32. Diagrama de secuencia Generar informe de Cuenta .....	46
Figura 33. Diagrama de secuencia para el análisis de la red .....	47
Figura 34. Diagrama de secuencia del caso de uso recuperar información .....	48
Figura 35. Diagrama de secuencia obtener clústeres usuarios .....	49
Figura 36. Diagrama de secuencia obtener clústeres bot .....	49
Figura 37. Diagrama de secuencia procesar información de red .....	50
Figura 38. Diagrama de secuencia generar informe red .....	51
Figura 39. Lista de tweets duplicados del usuario @AbrahamPriego96 .....	53
Figura 40. Proporciones del usuario @AbrahamPriego96 .....	53
Figura 41. Análisis de las principales entidades de @AbrahamPriego96 .....	54
Figura 42. Análisis de las principales categorías obtenidas @AbrahamPriego96 .....	54
Figura 43. Análisis de la frecuencia de actividad @AbrahamPriego96 .....	54
Figura 44. Análisis de la cuenta @AbrahamPriego96.....	55
Figura 45. Lista de tweets repetidos @blogAceiteOliva .....	56
Figura 46. Proporciones del usuario de @blogAceiteOliva .....	56
Figura 47. Análisis de las principales entidades de @blogAceiteOliva .....	57
Figura 48. Análisis de las principales categorías de @blogAceiteOliva .....	57
Figura 49. Análisis de la actividad de @blogAceiteOliva .....	57
Figura 50. Análisis de la cuenta de @blogAceiteOliva.....	58
Figura 51. Breve muestra del numeroso contenido duplicado por el usuario .....	58
Figura 52. Análisis de las proporciones obtenidas de todos los tweets extraídos del usuario .....	59
Figura 53. Entidades más destacadas del análisis del usuario. Fuerte implicación de otros usuarios en las publicaciones.....	59
Figura 54. Escasa categorización de los tweets del usuario debido al alto número de enlaces compartidos .....	59
Figura 55. Los tweets han podido ser analizados sin obtener un alto porcentaje de nulos.....	60
Figura 56. Análisis de la actividad del usuario con altos picos de actividad en determinados días y horas .....	60
Figura 57. Análisis de la cuenta del usuario. Se obtienen diferentes características para definir la conducta del usuario asociada a la cuenta.....	60
Figura 58. Análisis de tweets duplicados del @SergioRG92 sin muestra de tweets duplicados .....	61
Figura 59. Proporciones obtenidas de @SergioRG92.....	62
Figura 60. Análisis de las principales entidades de @SergioRG92 .....	62
Figura 61. Análisis de las principales categorías de @SergioRG92.....	62
Figura 62. Análisis del sentimiento de @SergioRG92.....	63
Figura 63. Análisis de la actividad de @SergioRG92.....	63
Figura 64. Análisis completo de la cuenta @SergioRG92 .....	63

Figura 65. Análisis de los tweets repetidos @Caiquet11 .....	64
Figura 66. Proporciones destacadas de @Caiquet11 .....	64
Figura 67. Análisis de las entidades principales de @Caiquet11.....	64
Figura 68. Análisis de las categorías principales de @Caiquet11 .....	65
Figura 69. Análisis de sentimiento de @Caiquet11.....	65
Figura 70. Análisis de la actividad de @Caiquet11 .....	65
Figura 71. Análisis general de la cuenta de @Caiquet11 .....	65
Figura 72. Lista de tweets repetidos de @TraedRuffles.....	66
Figura 73. Proporciones principales de @TraedRuffles .....	66
Figura 74. Análisis de las principales entidades de @TraedRuffles.....	67
Figura 75. Análisis de las principales categorías de @TraedRuffles .....	67
Figura 76. Análisis del sentimiento de @TraedRuffles .....	67
Figura 77. Análisis de la actividad de @TraedRuffles .....	68
Figura 78. Análisis general de la cuenta de @TraedRuffles .....	68
Figura 79. Lista de tweets semejantes entre los usuarios analizados .....	70
Figura 80. Comparación de proporciones entre los usuarios del análisis .....	70
Figura 81. Comparación de entidades entre los usuarios aportando gran valor al análisis.....	71
Figura 82. Comparación de categorías que aportan gran valor al análisis entre usuarios al obtener valores idénticos.....	71
Figura 83. Comparación de actividad entre los usuarios que arroja resultados exactos entre los usuarios .....	72
Figura 84. Comparación de porcentajes entre los usuarios analizados .....	73
Figura 85. Comparación de entidades entre los usuarios .....	73
Figura 86. Comparación de categorías entre los usuarios.....	74
Figura 87. Comparación de actividad entre usuarios que refleja una similitud en sus publicaciones .....	74
Figura 88. Mensaje si los usuarios no pertenecen a la misma red bot .....	75

# Capítulo 1

## INTRODUCCIÓN

El uso de redes sociales forma parte de lo cotidiano, presentando un medio de comunicación que permite el intercambio de información entre usuarios de forma inmediata. El 15 de enero de 2009, Twitter dio a conocer una noticia antes que los medios de comunicación [1], marcando el inicio de Twitter como medio de comunicación frente a los medios tradicionales. Esto es consecuencia directa de que prima la interacción con los usuarios [2] y la gran libertad que aporta a los usuarios para publicar cualquier tipo de información, ya sea en texto, video, foto o archivos.

Consecuencia de este potencial, Twitter es blanco de usuarios y redes de usuarios con pretensiones de propaganda para resaltar su contenido por encima del resto. Como Farhad Manjoo [3] dice, “si puedes lograr que algo sea popular en Twitter, tienes cobertura garantizada en todas partes”.

Algunos de estos usuarios son denominados bots. En el mundo de las redes sociales se conoce como bot a un programa que simula el comportamiento humano. Estas cuentas que publican información de forma automática tienen fines muy distintos. Desde hacer spam a cierto contenido a incrementar el número de seguidores de una determinada cuenta.

Actualmente existen bots cada vez más sofisticados que intentan eludir cualquier análisis, camuflando su perfil detrás de nombres coherentes, imagen de perfil y descripción aleatorias. Estos bots a menudo tienden a formar redes de bots con un mismo objetivo, como puede ser la promoción de un medio [2], o realizar spam político [4] [5].

En el mundo de las redes sociales se conoce como bot a un programa que simula el comportamiento humano. Estas cuentas que publican información de forma automática tienen fines muy distintos. Desde realizar spam a cierto contenido a incrementar el número de seguidores de una determinada cuenta.

Los bots han generado el 51,8% del tráfico en internet en el año 2016 según un estudio de la compañía británica de seguridad Imperva Incapsula [6]. Actualmente el 8,5% de las cuentas de Twitter son bots [7], llegando a un 15% según otros parámetros de análisis como indica la Universidad de California [8].

Muchos de estos bots tratan de interactuar con usuarios reales de la aplicación, pero a día de hoy no son capaces de realizar interacciones complejas con otros usuarios, aunque si interacciones con otros bots. Sin embargo, actualmente un gran número de estos bots son evidentes y poco sofisticados.

Existe un alto crecimiento en la proliferación de bots con fines informativos, de ayuda al consumidor, atención al cliente mediante chat, automatización de pedidos, entre otros [9]. Sin embargo, siguen dominando bots que pretenden influenciar en una opinión o decisión, o bots de spam para promocionar contenido [10].

El proyecto propone construir una aplicación web, TBotDetector, que analice la problemática de los bots y redes de bots en Twitter determinando si una cuenta puede ser considerada bot y si un conjunto de bots forman parte de la misma red. Para lograrlo se ha utilizado la ayuda de una herramienta Cloud llamada MeaningCloud que nos permite clasificar, categorizar y realizar un análisis de las entidades más destacadas y del sentimiento encontrado en los tweets de los usuarios, además de otras características relacionadas con la cuenta del usuario.

Destacar la existencia de otros servicios similares como IBM Watson [29] que permiten dar sentido al 80% de los datos que los expertos llaman “desestructurados” [37] cómo lo son los tweets de Twitter y así tomar mejores decisiones apoyadas en esa información.

## 1.1 OBJETIVOS

El objetivo del proyecto es crear una aplicación web para que identifique usuarios bots y redes de bots en Twitter. Se abordarán como principales los siguientes problemas:

- **Decidir si una cuenta es gestionada por un bot**  
La aplicación web evaluará el perfil del usuario y el contenido de su *timeline*. El objetivo es que el usuario sólo aporte el nombre de usuario que desea analizar y la aplicación pueda determinar si es un bot o no en función del análisis de su perfil y de su *timeline*.
- **Buscar conjuntos de bots que operen de forma conjunta**  
Se pretende distinguir redes de bots entre un conjunto de usuarios dados por el usuario.

Para resolver estos problemas, se plantean una serie de objetivos secundarios:

- **Obtener tweets de usuarios con las limitaciones que impone el API de Twitter**  
Debido a las restricciones del API de Twitter existen limitaciones de ventana temporal de 15 minutos (número de peticiones y de datos devueltos).
- **Gestionar heurísticas de detección de bots**  
Con el estudio previo de aplicaciones y trabajos que han detectado patrones de bots construyendo modelos de predicción, agrupar estas características para obtener un análisis efectivo.

## 1.2 ORGANIZACIÓN DEL DOCUMENTO

El presente documento está dividido en cuatro capítulos principales.

En primera instancia el “Estado del Arte” comprende el estudio de las tecnologías web empleadas, proporcionando los conocimientos necesarios para el correcto entendimiento y desarrollo de la aplicación.

En segundo lugar, se expondrá el diseño de la aplicación y su funcionalidad, resaltando las tecnologías empleadas para su creación.

Tas esta exposición, se presentarán ejemplos de informes generados con la aplicación. Abarca el ciclo completo de cada una de las funcionalidades principales desde la identificación de qué se quiere conseguir, pasando por cómo se ha conseguido desarrollar y cerrando con la efectividad final del caso que se propone.

Por último, evaluaremos cada uno de los objetivos planteados al principio de esta memoria realizando experimentos de los análisis obtenidos y redactando las líneas futuras a tener en cuenta para próximas implementaciones.

# Capítulo 2

## ESTADO DEL ARTE

Antes de comenzar el desarrollo de TBotDetector se ha realizado un estudio previo sobre el estado del arte.

Se comienza presentando Twitter como la red social escogida para el análisis. Una vez presentada la red social es necesario conocer qué tipos de bots existen actualmente y también cómo se coordinan para formar redes de bots. Por último, cuáles son las alternativas tecnológicas existentes y mecanismos que pueden ser de ayuda en el análisis de los perfiles de los usuarios.

Luego se elabora un análisis de las diferentes aplicaciones y trabajos que entre sus objetivos tratan de detectar cuentas bot. De este modo, es posible mejorar aquellos análisis hechos en otros estudios y conocer que aporta de nuevo TBotDetector a la detección de cuentas bot y redes de bot.

Una vez terminado este análisis, en la última sección se recogen las conclusiones de los análisis.

### 2.1 LA RED SOCIAL TWITTER

Twitter es una red social creada para la interacción entre usuarios mediante mensajes de texto breves. Creada en 2006 por los estudiantes de la Universidad de Cornell en Nueva York [11], Jack Dorsey, Biz Stone, Evan Williams y Noah Glass es una de las principales redes sociales a nivel mundial para el intercambio de información.

Produce contenido de manera instantánea recordando a los SMS (Short Message Service) debido a su limitación de 140 caracteres. Experimentó un elevado crecimiento gracias a la sencillez y semejanza que compartía con los nombrados SMS, aunque actualmente Twitter se encuentra en un pequeño estancamiento en el número de usuarios nuevos en la aplicación.

Existen dos tipos de relaciones entre usuarios de Twitter: aquellos a los que yo sigo (following ó siguiendo) y aquellos que siguen mi cuenta (follower ó seguidor), ver figura 1. Dentro de este marco pueden diferenciarse tres casos. El primero de ellos sería el seguimiento propio a otra cuenta sin ser seguido a continuación. El segundo, ser seguidor de un usuario que quiere estar al tanto de tus noticias, pero tú no decides seguirle. Y por último una relación recíproca de seguidor y seguido, definiendo así una relación mutua entre ambos.

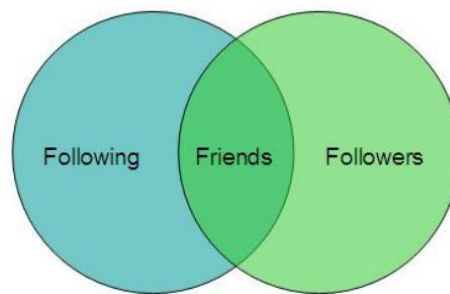


Figura 1. Relaciones posibles entre usuarios de Twitter

Para poder llevar un registro de nuestra actividad en Twitter, se guarda en nuestro perfil una lista ordenada cronológicamente con todos los tweets publicados o retuiteados de otras cuentas.

Existen otras mecánicas que fueron incluidas en Twitter a disposición de los usuarios cómo las menciones o la incorporación de etiquetas en los tweets que permite dotar de relevancia al texto o enfocarlo a un tema determinado y así categorizarlo.

La lista con todos los tweets publicados y retuiteados del usuario es una potente fuente de información que permite ser analizada con técnicas de análisis de textos para extraer información de ellos, además de otro tipo de información relacionada con la cuenta del usuario.

Para este proyecto, la fuente de datos utilizada es Twitter. De ella se van a extraer los tweets del usuario (limitado a 250) para realizar un análisis que permita identificar usuarios bot.

Se distinguen varias clases de bots en función de las características que presenten asociadas al perfil y a su contenido [34]. Estos perfiles son:

- **Bot seguidor:** con una alta proporción de seguidos respecto a seguidores. Este tipo de bots es creado para aumentar las cifras de seguidores sin importar el contenido de la cuenta. Suelen tener un

*timeline* de escaso valor informativo, arrojando análisis pobres respecto al contenido de la cuenta.

- **Bot de spam:** son bots programados para promocionar un hashtag o sitios web duplicando contenido. En este tipo de bots el *timeline* si tiene relevancia y por eso se obtienen análisis potentes de su contenido.
- **Bot retuits:** suelen pertenecer a cuentas influyentes capaces de posicionar un determinado contenido en la red social. Comparten rasgos con los bots spam, ya que suelen compartir enlaces o determinados hashtags.

En relación con los tres tipos nombrados anteriormente, los patrones temporales también permiten identificar perfiles bot. Los usuarios tuitean con menor frecuencia que los bot. Publican su contenido sin un patrón de horario fijo.

## 2.2 APLICACIONES PARA LA DETECCIÓN DE BOTS

Existen diferentes aplicaciones que se centran en la detección de perfiles bot. A continuación, se destacan las principales características de cada una de ellas.

**Bot Or Not** [12] [13] fue Desarrollado por la universidad del Sur de California en colaboración con la universidad de Indiana [14], es capaz de detectar en base a más de 1000 parámetros la veracidad o no de ese usuario [15]. Elementos como tipo de posts, cantidad de seguidores y periodicidad de los tweets son medidos y verificados por la app, para llegar a un veredicto con respecto a determinado perfil de Twitter. Para poder utilizar esta APP, ver figura 2, es necesario registrarse con el usuario de Twitter y dar permisos para que pueda acceder a tu cuenta.

Las conclusiones arrojadas por este estudio permiten identificar clústeres de usuarios en función de los datos obtenidos en los más de 1000 parámetros analizados, proporcionando un modelo fiable para nuestro proyecto.

La herramienta **Fake Follower Check** [16] es capaz de analizar la lista de seguidores de una determinada cuenta y mostrar el porcentaje de perfiles que no tienen ningún tipo de actividad, los que son falsos y aquellos que son reales. Existe la posibilidad de extraer más información de los perfiles como por ejemplo el idioma que emplean los usuarios, si se contrata uno de los planes que ofrecen a un determinado precio.

**Twitter Audit** [17] se encarga de analizar cualquier cuenta tomando como referencia 5000 seguidores del usuario y le asigna una puntuación según el número de tweets, la fecha del último tweet y la proporción de seguidores. Una vez recogidos estos datos asigna un porcentaje que señala la veracidad de la cuenta.



**Tweet Cleaner** [15] es una herramienta que proporciona una clasificación de usuarios atendiendo a los siguientes criterios. Detecta bots spam con un alto uso de links o tweets duplicados. Usuarios inactivos que no hayan actualizado su perfil en un mes o posean menos de diez tweets. También es capaz de detectar contenido no original si más del 70% de los tweets son retuits y perfiles bots si más del 90% de los tweets proceden de otra fuente de información.

**TheBotBuster** [35] es una web que permite realizar un análisis de cualquier usuario de Twitter que se desee. La aplicación permite categorizar el usuario en tres escenarios. El primero de ellos, las cuentas humanas, pasando a las cuentas administradas por seres humanos u organizaciones y por último las cuentas bot.

Mediante una escala permite asignar una nota en base a todos los análisis que ha realizado. Los análisis abarcan aspectos de la cuenta del usuario, como de la interacción con otros usuarios de Twitter.

Existen también trabajos de distintas universidades que han servido de apoyo para la creación de este proyecto. Entre los más destacables se encuentran los siguientes:

**Herramienta automática para la detección de bots**, trata de un estudio que intenta distinguir cuentas bot dada la información que los usuarios comparten en la red social [18]. Se centra en tres hipótesis de partida cuya validez permitirá diferencias las cuentas bot de las cuentas de humanos. Estas hipótesis son:

Las cuentas controladas por bots escriben tweets de forma regular, mientras que los humanos cuentan con intervalos de tiempo sin tuitear.

Los tuits de las cuentas bot contienen un alto porcentaje de enlaces frente a las cuentas controladas por humanos.

Los tuits de las cuentas que son bot utilizan menos variedad de contenido que los usuarios humanos.

**Herramienta para el análisis de la interacción, identificación de patrones y clasificación de usuarios en humanos, y bots de la red Twitter.**

El proyecto [19] plantea una solución para identificar cuantos de los usuarios activos son humanos y si es posible detectar de forma automática si un usuario es bot.

El análisis se realiza a partir de una muestra de datos extraídos de la red Twitter con la intención de crear un modelo que, en base al contenido generado por los usuarios, su comportamiento y relaciones con el resto de perfiles, pueda aportar una clasificación de forma automática entre un bot y un humano.

## 2.3 CONCLUSIONES

Se observa para cada una de las herramientas mencionadas anteriormente en el apartado 2.3, que existe una base en el análisis de usuarios en la red social Twitter. Las aplicaciones y los estudios revelan que los sistemas de detección basados en características pueden revelar patrones de comportamiento interesantes para clasificar a los usuarios entre bots y humanos. Los metadatos del usuario se consideran entre las características más predictivas, permitiendo inferir reglas que permitan tomar decisiones para clasificar el usuario. La ventaja de centrarse en los patrones de comportamiento es que estos pueden ser codificados para aprender la tendencia de un bot y un humano.

Sin embargo, los bots cada vez son más sofisticados [14] intentando mezclar conductas bot con humanas ya que introducen nuevas dimensiones para emular, además del contenido, incluyendo la red social, la actividad temporal, los patrones de difusión y la expresión del sentimiento. La detección de estos bots es muy complicada para los sistemas basados en características si consiguen copiar el comportamiento de un usuario no bot.

Gracias a estos estudios es posible definir las técnicas para detectar bots y los requisitos que se necesitan implementar para alcanzar los objetivos. Estos requisitos serán descritos en la sección 3 de esta memoria.

# Capítulo 3

## REQUISITOS

Como se ha podido ver en el apartado 2.3 de la presente memoria, existen aplicaciones que tratan de detectar bots en Twitter basándose en un análisis de las características que permiten distinguir entre bot y humano. Sin embargo, no se realiza una distinción entre los tipos de bots que existen en la red social y tampoco la posibilidad de detectar grupos con determinados patrones de conducta, es decir, redes de bots en Twitter.

En relación a esta última novedad, los estudios actuales [30] [31] indican que hay un incremento de la manipulación en las cuentas de redes sociales con fines propagandísticos lo que provoca la aparición de redes de bots creadas para un mismo propósito.

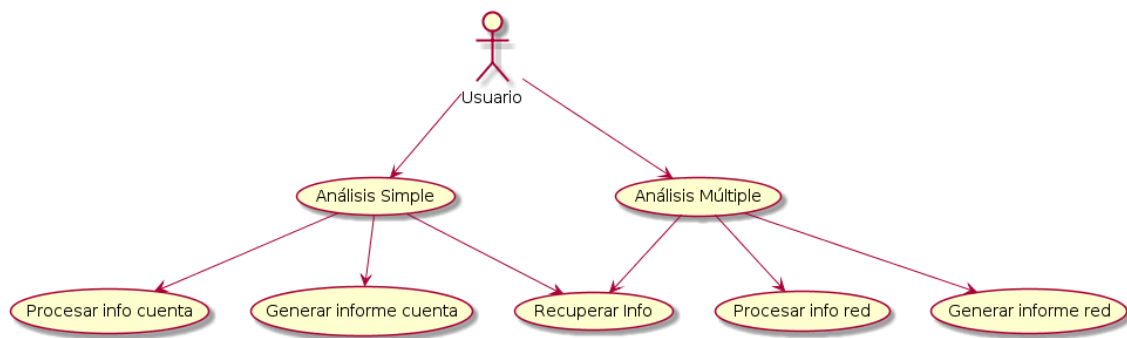
TBotDetector es una aplicación web. Se basa en un análisis estadístico y de lenguaje natural de los tweets de los usuarios. Mucha de esta funcionalidad ha sido tomada de MeaningCloud, una herramienta que proporciona servicios en la nube para procesamiento de textos.

Con esta funcionalidad, TBotDetector quiere categorizar los tipos de bots que pueden aparecer en Twitter, buscando evidencias que apoyen o desmientan si un usuario es un bot y elaborando informes prospectivos.

El mismo fin se persigue con las redes bots, aunque en este caso se trata de averiguar si dos presuntos bots operan de forma conjunta.

### 3.1 REQUISITOS FUNCIONALES

La aplicación web desarrollada tiene como objetivo el análisis de perfiles en la red Social Twitter para detectar bots y redes de bots entre los usuarios. Para ello proporciona las siguientes funcionalidades. En la figura 2 se muestra el escenario definido identificando los casos de uso de la aplicación.



**Figura 2. Casos de uso de la aplicación**

Estas funcionalidades representan los dos análisis realizados por TBotDetector. El caso de uso análisis simple se centra en un único usuario que se desee estudiar, mientras que el caso de uso análisis múltiple permite detectar redes de bots.

A su vez, cada uno de estos casos de uso son complementados con otros. Para ambos análisis es necesario recuperar la información de la cuenta y contenido del perfil o perfiles de Twitter, ya que es la fuente de datos que se emplea para el análisis.

Es el primer paso para el estudio. El segundo paso permite procesar la información recogida para adaptarla a las necesidades de cada análisis. Por último, estos análisis que se han creado necesitan aportar valor al estudio, es decir, se generan los informes de cuenta para el análisis simple y el informe de red para el análisis múltiple

### 3.1.1 ANÁLISIS SIMPLE

Busca evidencias para determinar si es un bot. Se da la posibilidad al usuario de escoger el perfil de Twitter que se desea analizar. Esta propuesta se centra en dos factores fundamentales para construir un modelo de clasificación.

Para obtener estos factores se ha recurrido a los estudios y aplicaciones analizados en el capítulo 2 de la memoria. Se han adaptado los análisis vistos en estos estudios y los obtenidos a través de MeaningCloud.

#### **Factores relacionados con el perfil del Usuario**

Se estudian las características referentes al comportamiento del usuario dentro de Twitter.

- Perfil verificado en Twitter
- Proporción de Seguidores
- Reputación de la cuenta ( $\text{n}^{\circ}$  seguidores / [ $\text{n}^{\circ}$  seguidores /  $\text{n}^{\circ}$  seguidos])
- Comprobar si contiene tweets duplicados
- Comprobar si la foto de perfil no es la que te asignan por defecto
- Comprobar si posee geolocalización
- Distribución de los tweets por semana
- Distribución de los tweets por día

#### **Factores basados en el contenido de los Tweets (Clústeres del timeline)**

Se estudia el contenido publicado por el usuario. Estas técnicas aportan gran valor para el análisis y son las que permiten encontrar patrones en el usuario para el análisis de entidades, la categorización de textos y el análisis de sentimiento.

Estos factores serán obtenidos a través de los datos proporcionados por el análisis que ofrece MeaningCloud descrito en la sección 2.8 de este proyecto. Permiten ampliar los análisis vistos en los estudios de otros trabajos que también han sido tenidos en cuenta para la creación de estos factores [18] [19].

Con todo ello podemos obtener los siguientes factores:

- Proporción de menciones y las más destacadas
- Proporción de URL y las más destacados
- Proporción de Hashtags y los más destacados
- Proporción de retweets
- Número de Tweets repetidos
- Alta o baja clasificación de tweets
- Diversidad de temas o centrado en uno concreto
- Entidades destacadas y número de apariciones
- Valor del análisis de sentimiento.

Cada uno de estos factores arrojará un resultado individual que juntos proporcionarán la información necesaria para tomar una decisión. Con este análisis se pretende ampliar el simple análisis de detectar si un perfil es bot o no y poder extraer más información de las características que definen al bot.

Los siguientes diagramas de actividad muestran los algoritmos principales para el análisis simple. La figura 3 detalla el proceso de negocio principal para obtener la información sobre la cuenta del usuario introducido en la aplicación.

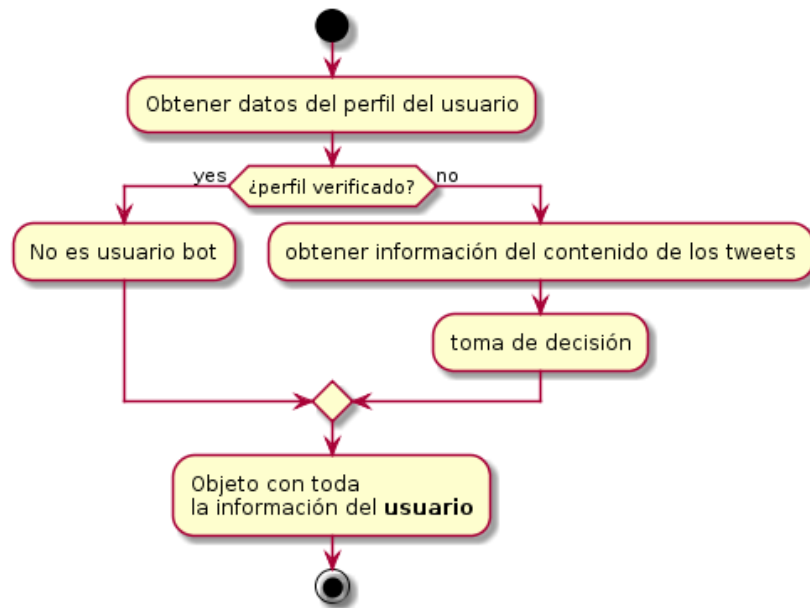


Figura 3. Algoritmo principal para la detección de bots en el análisis simple

La obtención de la información de los tweets es un proceso de recopilación de datos proporcionados por la librería MeaningCloud. El proceso es el mostrado en la figura 4.

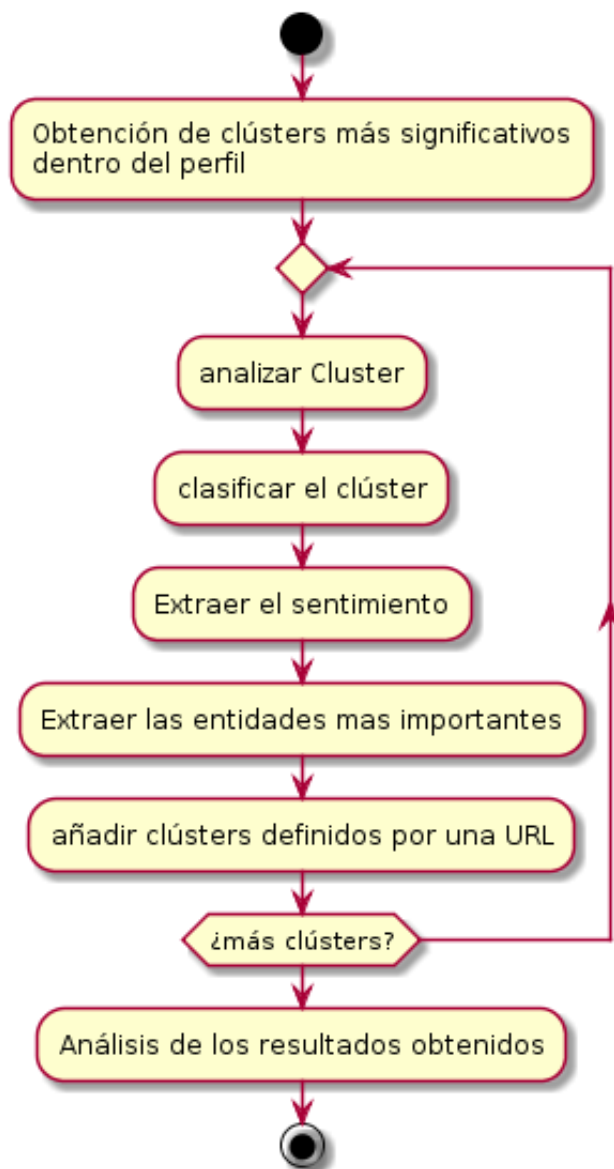


Figura 4. Extracción de clústeres más relevantes del usuario en función de la longitud o URL como similitud

Por último, los algoritmos que permiten analizar cada una de las características que aportaran valor al análisis. Cada una de las figuras que se muestran a continuación representa un análisis.

En la figura 5 se puede apreciar la toma de decisión respecto los datos obtenidos del perfil del usuario.

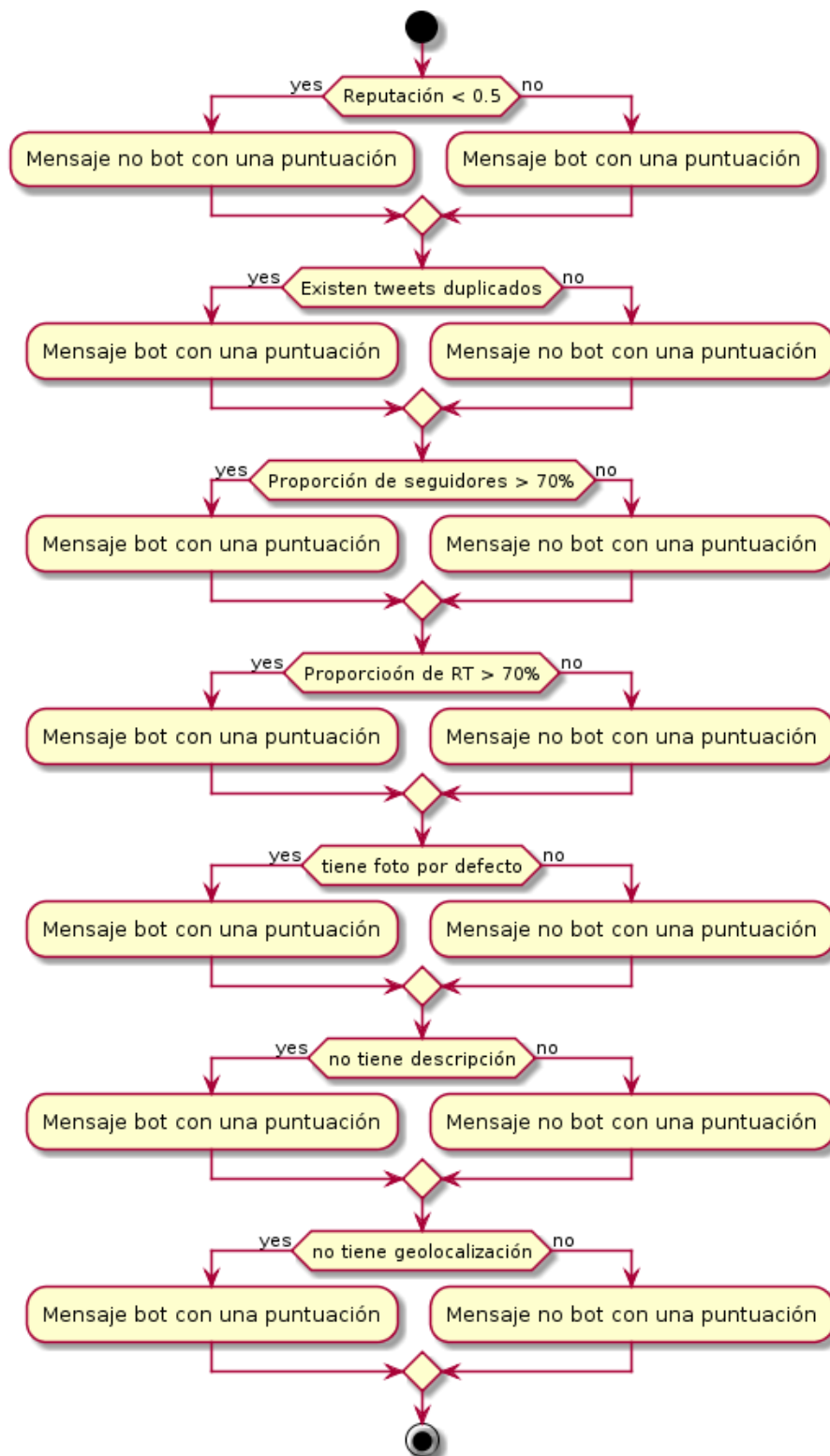


Figura 5. Toma de decisión en función de los datos de la cuenta del usuario



La figura 6 representa características globales obtenidas del contenido de los tweets. Se toman decisiones respecto a los porcentajes obtenidos.

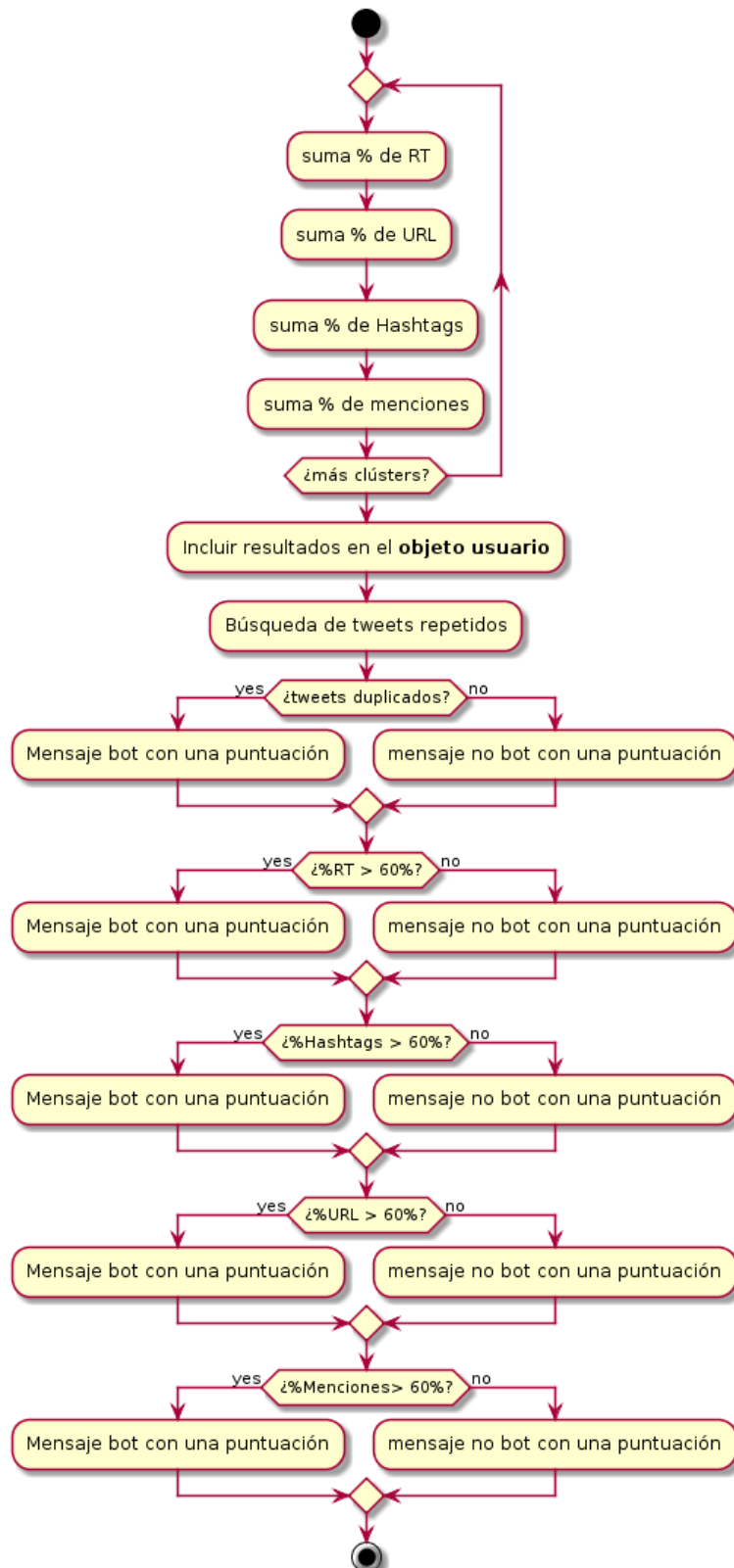


Figura 6. Toma de decisión en función de determinadas características del usuario

La figura 7 realiza un análisis respecto a las categorías en las que se han podido dividir los tweets.

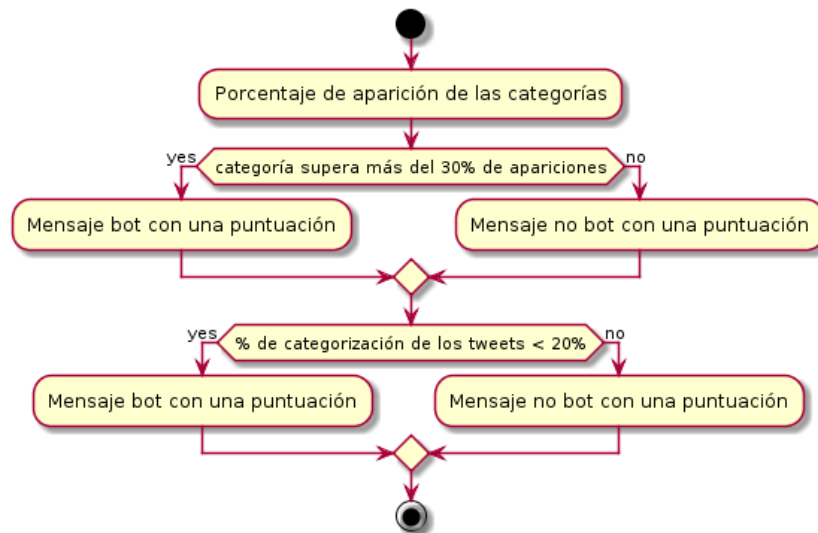


Figura 7. Análisis de las categorías más destacadas del usuario mediante comparación de proporciones

La figura 8 obtiene el análisis respecto a los porcentajes obtenidos en el análisis de entidades.

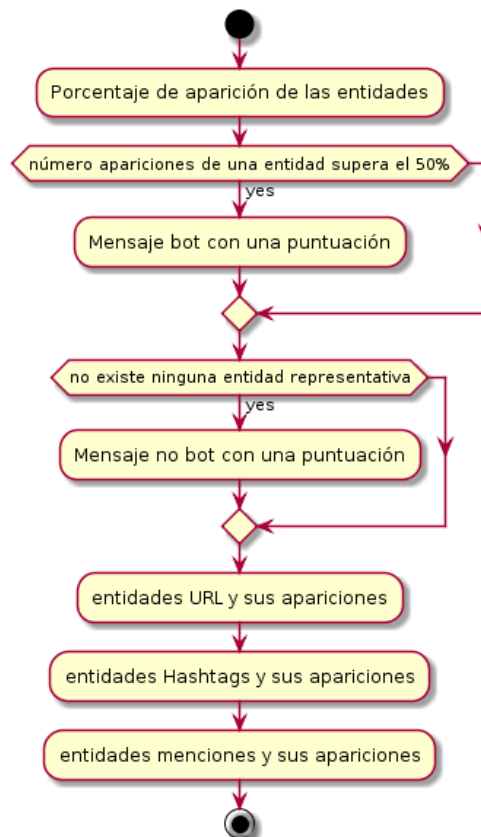


Figura 8. Análisis de entidades más representativas y número de apariciones

Por último, la figura 9 representa el análisis obtenido en la obtención de sentimiento de los tweets. Este análisis reveló que aquellos usuarios bot, obtenían porcentajes elevados de polaridad nula en los análisis.

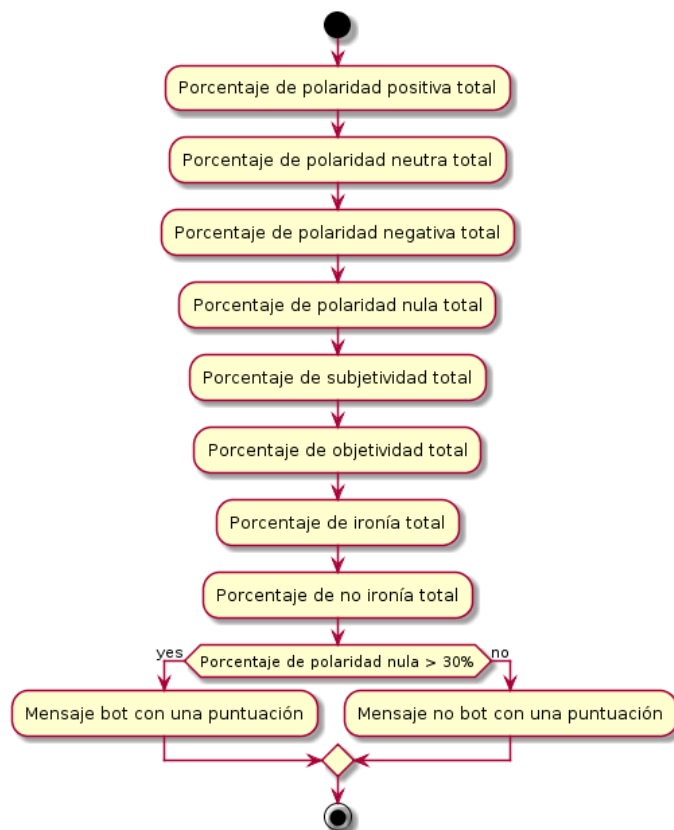


Figura 9. Análisis de sentimiento y toma de decisión para el análisis en función de la polaridad obtenida

### 3.1.2 ANÁLISIS MÚLTIPLE

Teniendo como base las técnicas para detectar bots y estudios que hacen referencia a redes de bots [2] [20], se considera para este análisis unas características similares a las empleadas para la detección de bots.

Este análisis permite identificar redes de bots formadas por cuentas que tienen un perfil similar. Se da la posibilidad de importar un archivo con los nombres de la posible red de bots. Recaltar que el primero de esta lista será el usuario bot con el que se van a comparar el resto de cuentas. Los requisitos a implementar son:

- Obtención de los clústeres comunes en cada comparación. Lista con los tweets más repetidos, las entidades más nombradas de cada clúster y las categorías en las que han sido posible categorizar los tweets.
- De los clústeres obtenidos anteriormente, recoger la distribución en el tiempo de los tweets para establecer patrones temporales entre ambos perfiles.

Al igual que para el análisis simple, se describe en el primer diagrama la funcionalidad principal del algoritmo para la obtención de redes de bots. Ver figura 10.

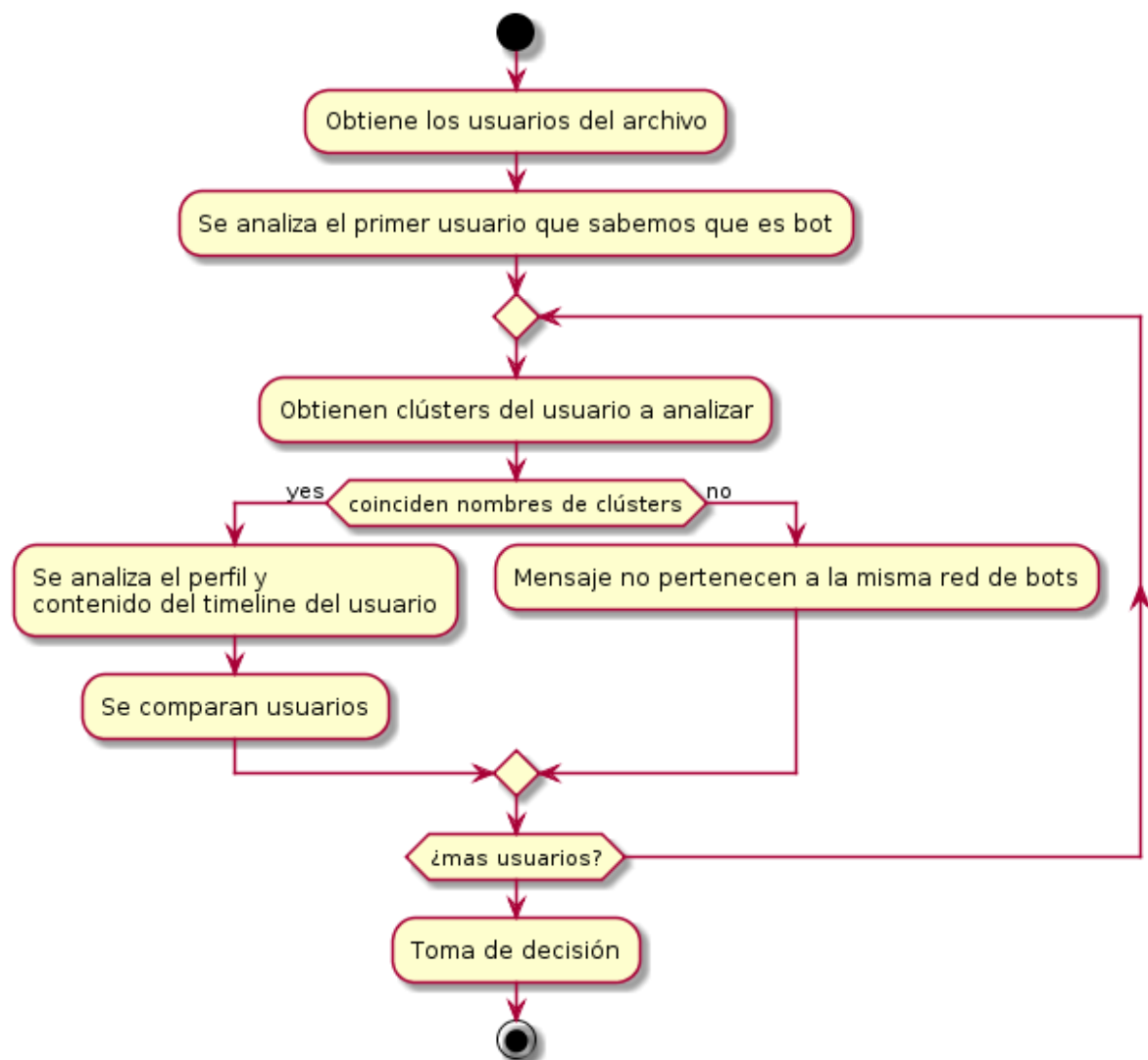


Figura 10. Algoritmo principal para la obtención de redes bot

La siguiente funcionalidad a destacar se centra en la obtención de la información del usuario que se desea comparar, a través de MeaningCloud como se ha mostrado en la figura 11 del análisis simple. Una vez obtenida esta información se comparan ambos usuarios en diferentes algoritmos.

Estos algoritmos estudian una determinada característica en comparación con el usuario bot.

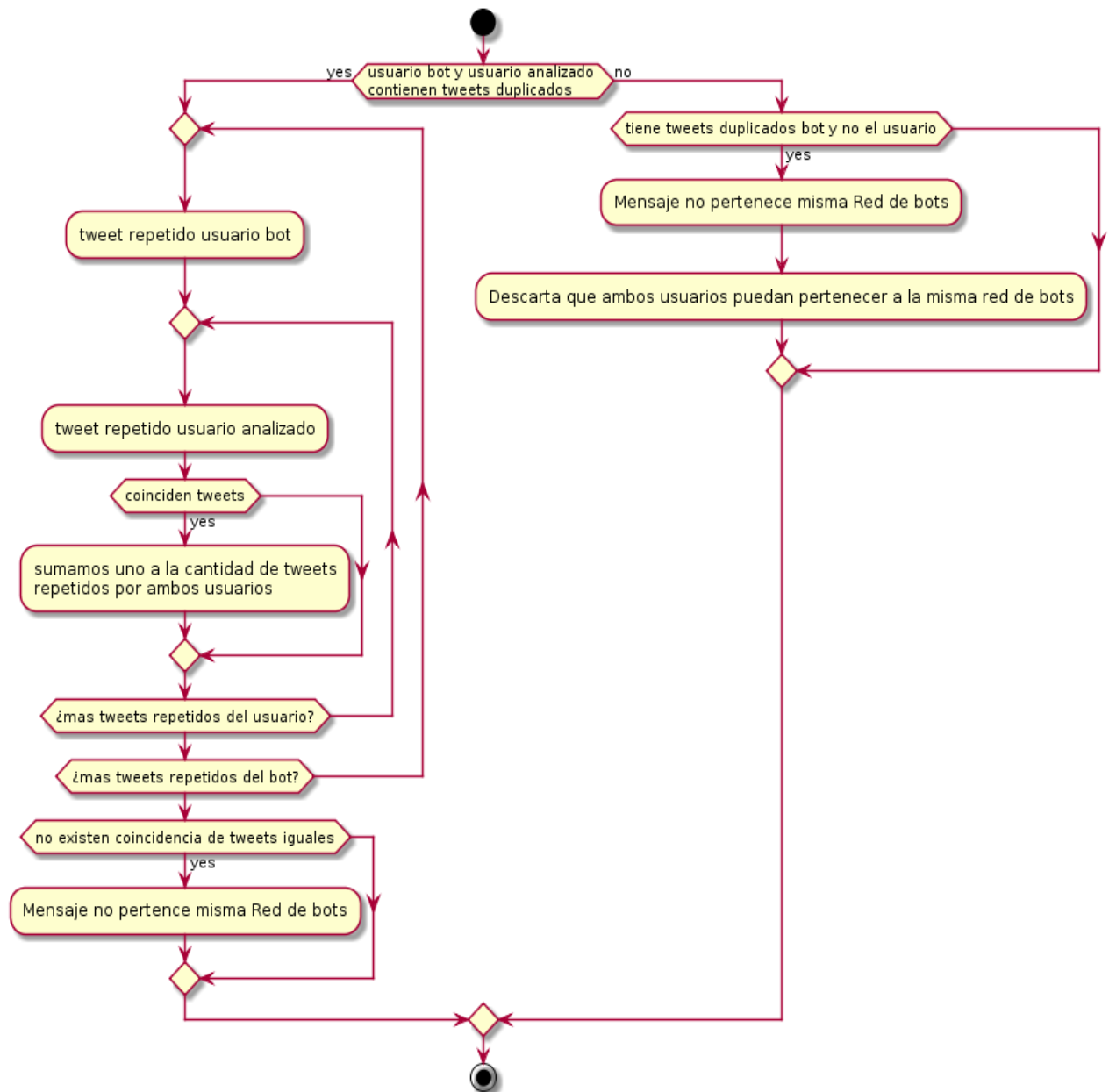


Figura 11. Análisis de tweets duplicados y comparación entre el usuario bot y el analizado

La figura 12 muestra las comprobaciones que se realizan una vez obtenidos los clústeres que coinciden entre los usuarios. Se estudian los duplicados de los tweets y la diferencia de proporción entre diferentes análisis.

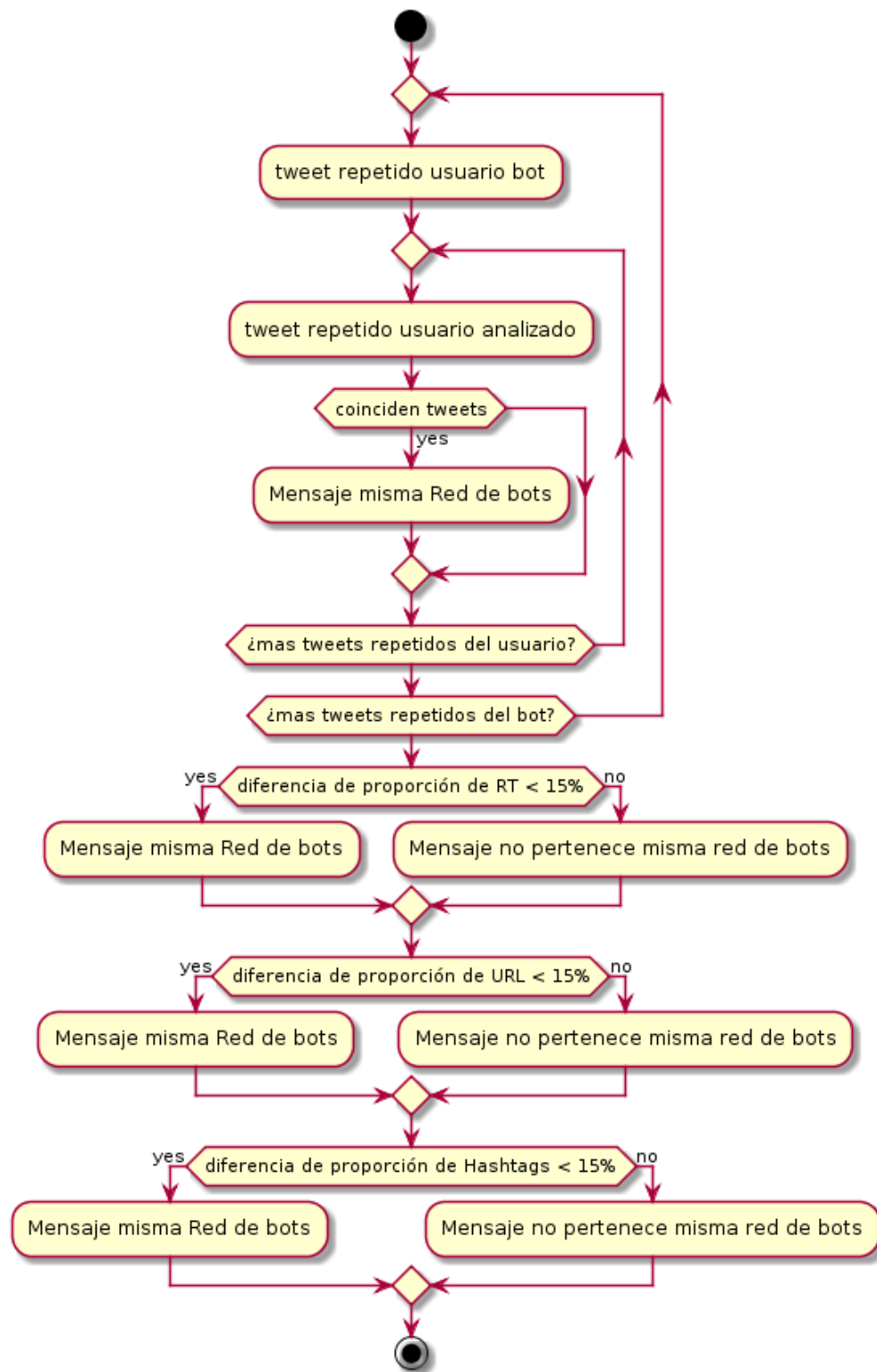


Figura 12. Comprobaciones de la información de los clústeres para tomar una decisión

El análisis continúa con la figura 13. Representa la comparación entre entidades de ambos usuarios en busca de coincidencias que puedan afirmar que tienen a ser bot representantes de la misma red.

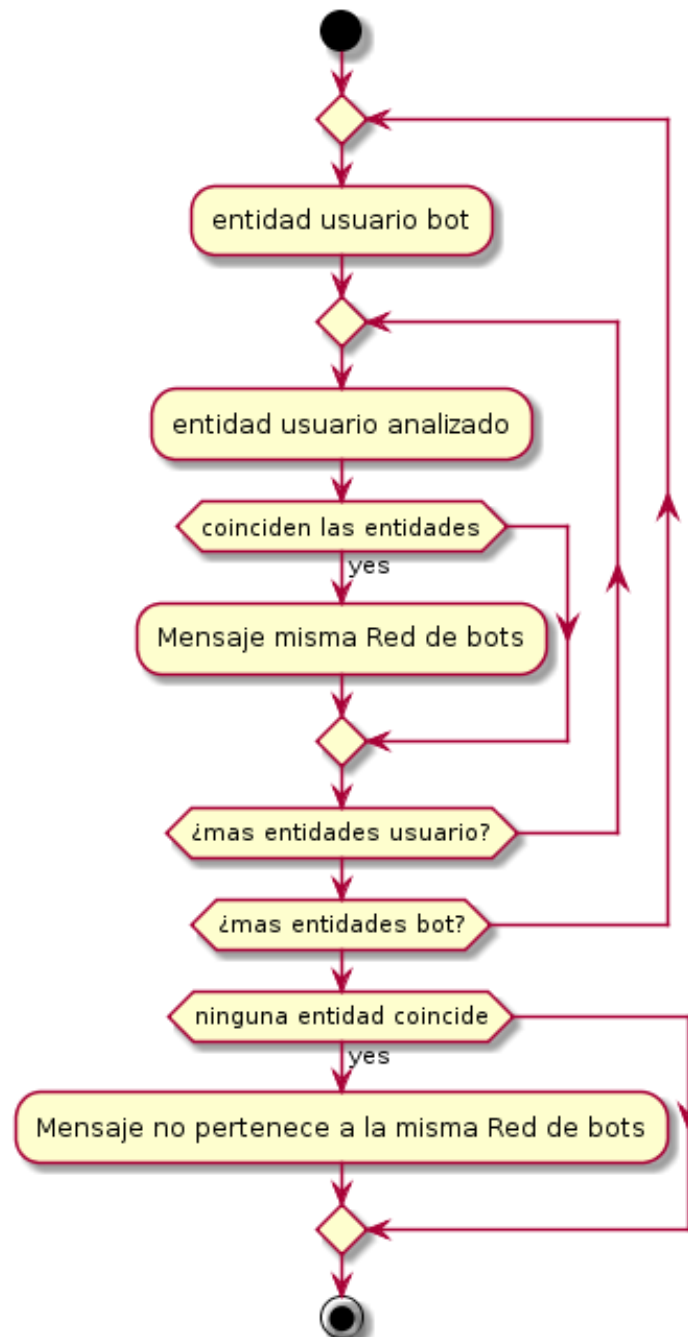


Figura 13. Comparación de entidades para el análisis de redes bot

La figura 14 representa la comparación de categorías entra ambos usuarios. Como ocurría en el análisis anterior, si las categorías coinciden, tienen a formar parte de la misma red de bots.

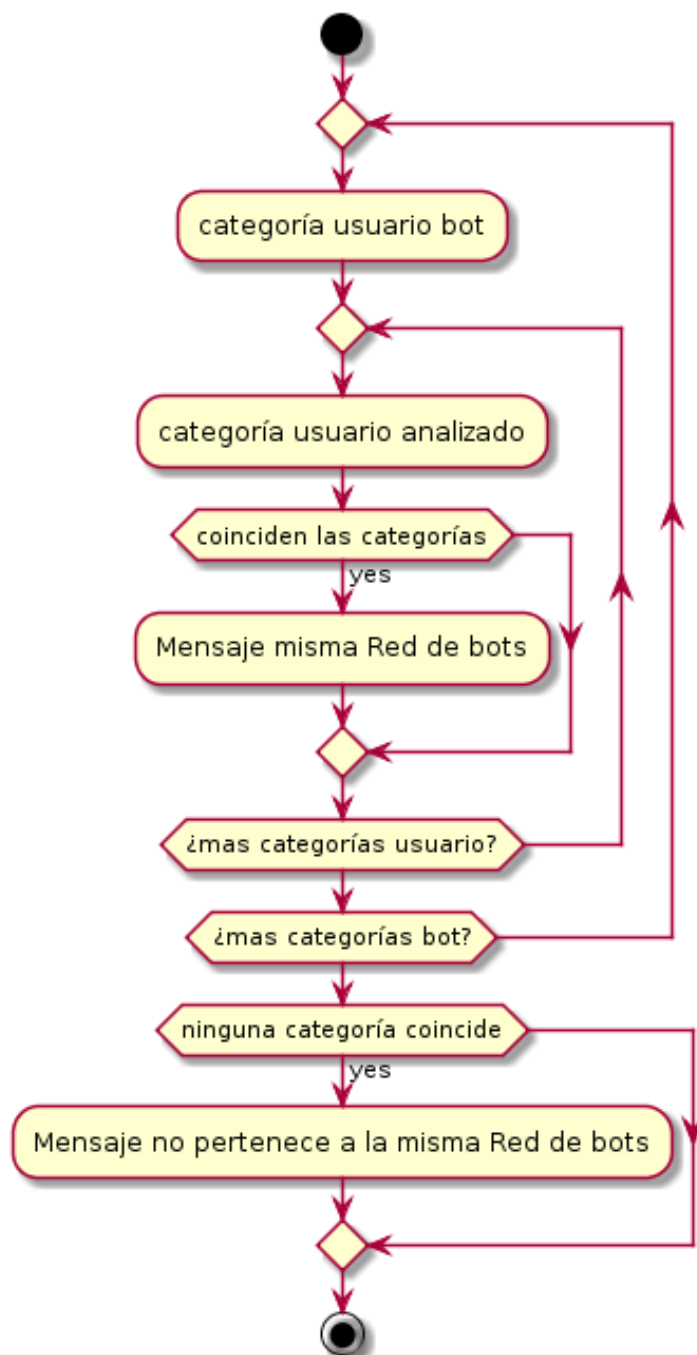


Figura 14. Comparación de categorías para el análisis de redes bot



Por último, la figura 15 representa las dos últimas comparaciones del análisis. Se comprueba si existen picos de actividad en determinados días y en determinadas horas que representen la posibilidad de pertenecer a una misma red de bots.

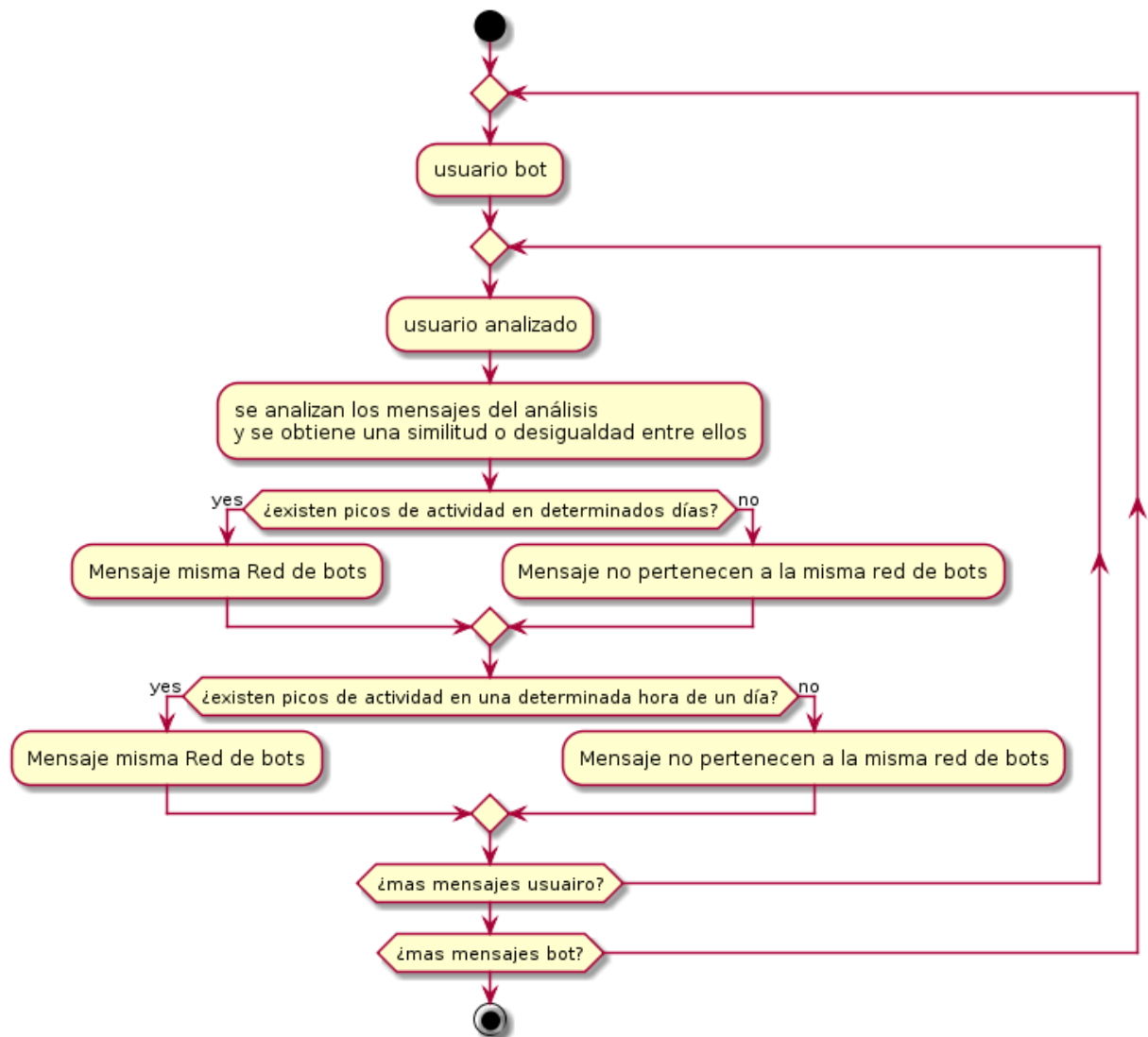


Figura 15. Análisis de la actividad de los usuarios y toma de decisión en base a los resultados obtenidos en la comparación

### 3.2 REQUISITOS NO FUNCIONALES

A continuación, se definen los requisitos no funcionales de la aplicación web que da acceso a esta funcionalidad:

- Interacción con el usuario por medio de un navegador web ya sea en escritorio, Tablet o Smartphone.
- Funcionamiento en versiones actuales de los navegadores. El uso de navegadores no actuales puede comprometer la visualización de los datos.
- Es necesaria una conexión a internet para la ejecución de la aplicación web.

Las pruebas se han realizado en un pc de sobremesa y un portátil en navegadores Chrome, Firefox y Edge.

# Capítulo 4

## DISEÑO

En este capítulo se presenta primero la interfaz gráfica haciendo hincapié en las tecnologías usadas y la forma en cómo se presenta en distintos dispositivos. A continuación, se presenta la interacción entre las vistas para los distintos análisis y el diseño escogido para la interacción del usuario. Por último, se da a conocer el servicio MeaningCloud implementado para la obtención de valor en la información recogida de Twitter.

TBotDetector es una aplicación web que se basa en diferentes tecnologías para lograr una mejor interacción con el usuario. La primera de esas tecnologías es Thymeleaf [32] [33]. Motor de vistas que proporciona una distribución de vistas estática para construir el layout principal de la aplicación donde se incluirán el resto de vistas en función las peticiones enviadas al controlador que será el encargado de redirigir el contenido (vista dinámica) que será mostrada al usuario.

La visualización de los componentes está a cargo de la librería Material Design [23] creada por Google permitiéndonos agregar un diseño a la aplicación. Permite optimizar el uso de dispositivos cruzados y así tener una experiencia de usuario agradable en cualquier navegador o dispositivo.

### 4.1 INTERFAZ GRÁFICA

El usuario interactúa mediante una aplicación web. Es un diseño responsivo, lo que permite adaptarse a cualquier dispositivo ya sea pc, portátil, Tablet o Smartphone. En la figura 16 se visualiza el estado de la página web visto desde un pc o portátil.

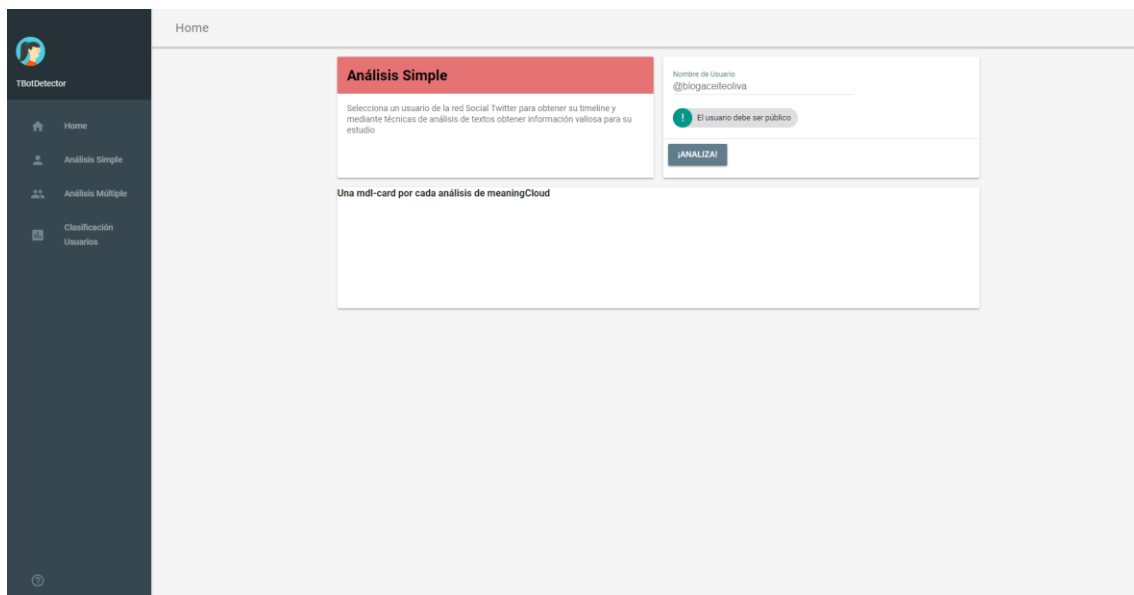


Figura 16. Vista inicial del análisis simple

El diseño cambia cuando el dispositivo usado es un móvil o una Tablet. Se oculta el menú lateral para una mayor visibilidad de la parte principal de la aplicación. La figura 18 representa la interfaz de la aplicación para empezar el análisis simple en un Smartphone.

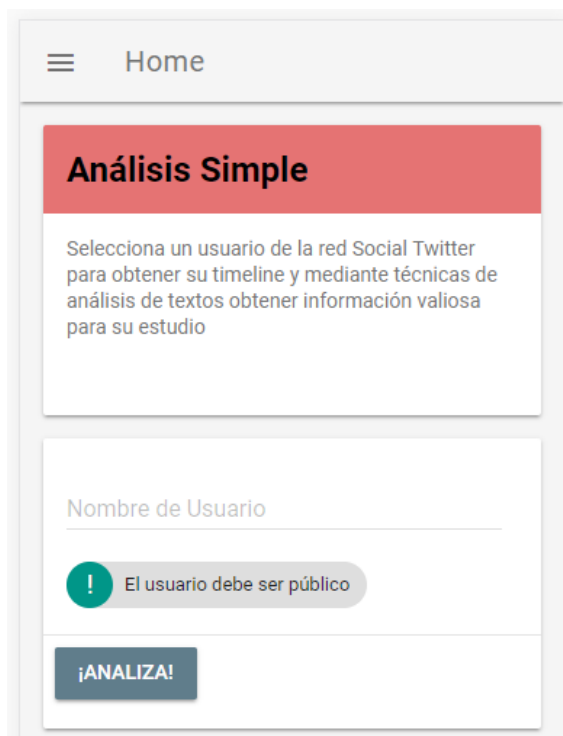


Figura 18. Vista del análisis simple desde Smartphone

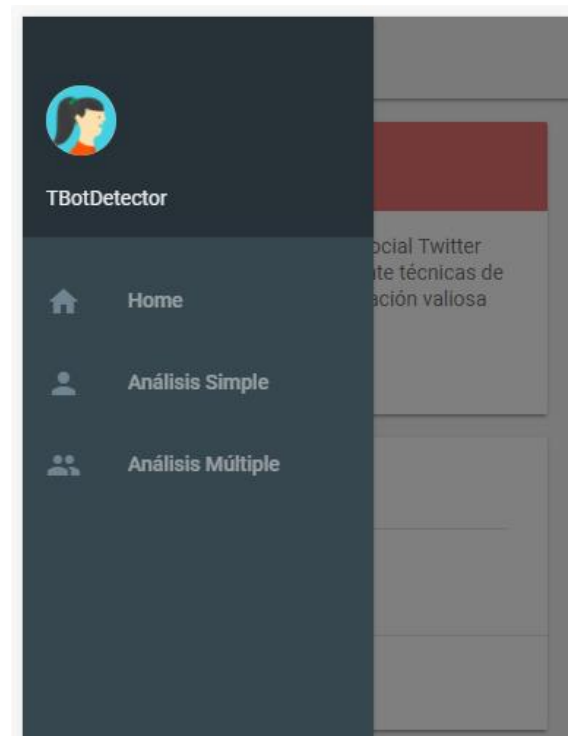


Figura 17. Vista del menú desplegable desde Smartphone

La figura 17 representa el menú desplegable visto desde un Smartphone. Mientras que, la figura 19 representa la interfaz para comenzar el análisis simple en Tablet.

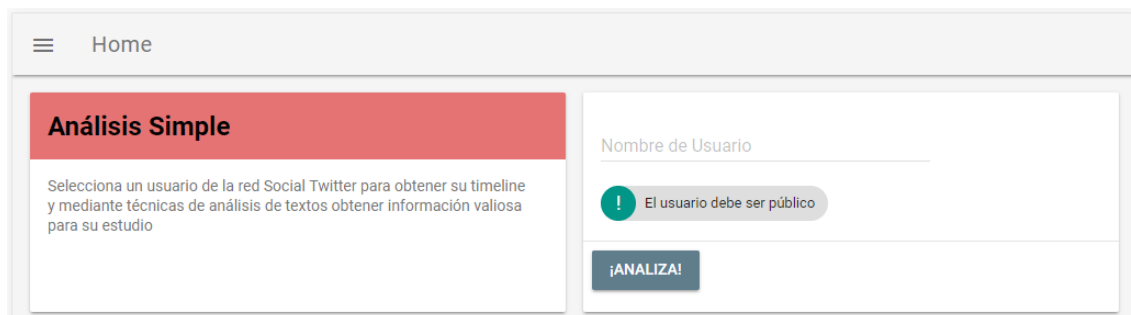
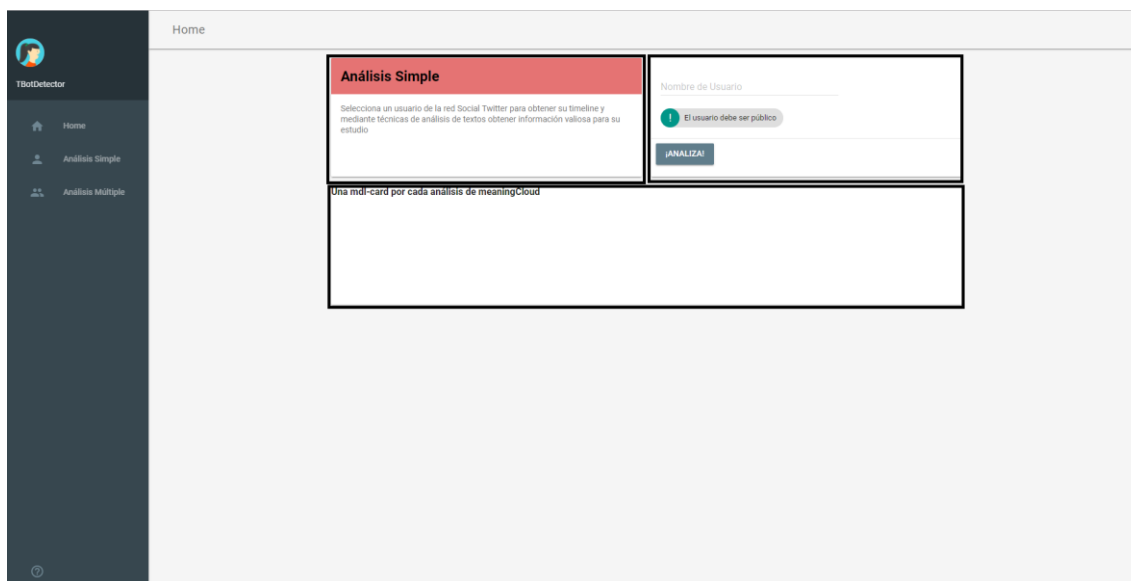


Figura 19. Vista del análisis simple visto en formato Tablet

En la figura 20 se puede distinguir los dos layout que forman la aplicación. La parte dinámica de la página web está representada por un rectángulo. Mientras que el resto de la página, menú lateral y encabezado forman la parte estática. Esto quiere decir que la aplicación mantendrá un único HTML en todo el funcionamiento.

El contenido de este HTML variará en función de la opción seleccionada en el menú lateral o por los botones que se muestran en la página que se muestra en la primera carga de la aplicación.

La interacción entre cada una de las vistas de la aplicación es sencilla. Basta con el menú lateral que ofrece para navegar entre una u otra. También es posible acceder a las páginas que dan acceso a ambos análisis desde la página de inicio de la aplicación.



**Figura 20. Vista principal del análisis múltiple**

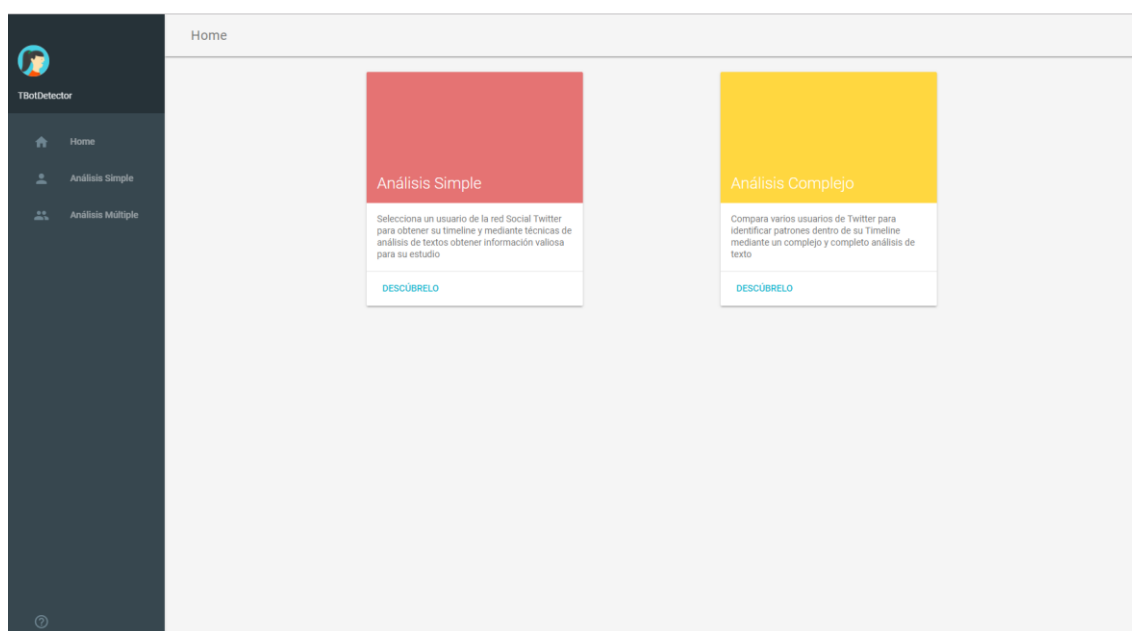
En la figura 20 vista anteriormente, se puede observar cómo es posible acceder a ambos análisis seleccionando la opción en el menú lateral o en el recuadro que describe el análisis.

Una vez se carga una de las dos páginas se realiza el análisis y los resultados serán vistos en el recuadro que se situará justo debajo de los dos recuadros iniciales. Ver figura 21. Esta funcionalidad estará disponible en las siguientes versiones de TBotDetector, ya que actualmente, los análisis se visualizan por consola.

Si un análisis ha terminado, es posible redirigirse a cualquiera de las otras vistas de la aplicación, ya sea la página inicial o el otro análisis. Destacar que es posible volver a efectuar el análisis cargando nuevos usuarios.

El diseño de las vistas se ha creado para que el usuario no tenga que navegar entre pantallas y pueda obtener la información que necesita para realizar un análisis y la información que desea ver del análisis en la misma página.

La estructura de cada una de las vistas es idéntica en ambos casos de estudio. La estructura está definida en tres bloques diferenciados 21.



**Figura 21. Vista principal de la aplicación**

En una primera carga de la página se visualizarán los dos primeros bloques. El bloque situado a la izquierda de la página es una breve explicación de qué se va hacer en ese análisis y qué debe hacer el usuario.

El segundo de los bloques, situado a la derecha, permite la importación de un archivo que contendrá un nombre de usuario, o un conjunto de nombres, en función de la página en la que nos encontremos. En ambos análisis se dispondrá de un botón para comenzar el proceso.

Por último, el bloque del análisis que informa de si el perfil es bot o no será mostrado una vez el usuario haya elegido el archivo a importar y ejecutado el análisis.

#### 4.1.1 MEANINGCLOUD

MeaningCloud [24] es una herramienta Cloud para la analítica de textos que permite extraer conocimiento de cualquier texto. De forma sencilla, potente y asequible permite extraer información de gran valor para análisis de contenido no estructurado. MeaningCloud no resuelve el problema de determinar si un usuario es un bot, pero proporciona funcionalidad que ayuda a lograrlo.

Como tecnología Cloud presenta una serie de ventajas e inconvenientes. Entre las ventajas, destacar la posibilidad de acceso desde cualquier dispositivo o lugar. Los servicios permiten pagar por sólo aquello que utilizas, teniendo siempre la tecnología actualizada. No requiere del uso de aplicaciones de terceros y los datos se

encuentran siempre disponibles. Tiene el inconveniente de que es necesario tener una buena conectividad al no tener acceso a un servidor físico.

Provee un API en la nube para desarrolladores. Para este proyecto vamos a centrarnos en dos soluciones que proponen mediante su API.

La primera de estas soluciones es el **Análisis de redes Sociales**. [25] Aporta solución al problema que conlleva analizar información no estructurada de redes sociales ya que la inmensa mayoría se presenta en este formato.

Esto permite identificar temas y tendencias al igual que entender automáticamente la estructura y el significado de noticias en medios sociales, más allá de una simple agregación de apariciones o menciones, extrayendo valor de los elementos más representativos con mayor precisión y velocidad que el ser humano.

La segunda trata la **analítica de textos**. [26] Han creado soporte para extraer automáticamente información de alto valor de los textos. Aporta un gran valor en la monitorización y análisis de medios, para analizar la información generada por los usuarios.

Tanto el análisis de redes sociales como la analítica de textos son aplicables a este proyecto. Permiten obtener características y patrones de los tweets. Estos patrones permiten identificar los diferentes tipos de usuarios que se encuentran en la red social. La figura 22 muestra algunas de las características que se pueden obtener del texto analizado.



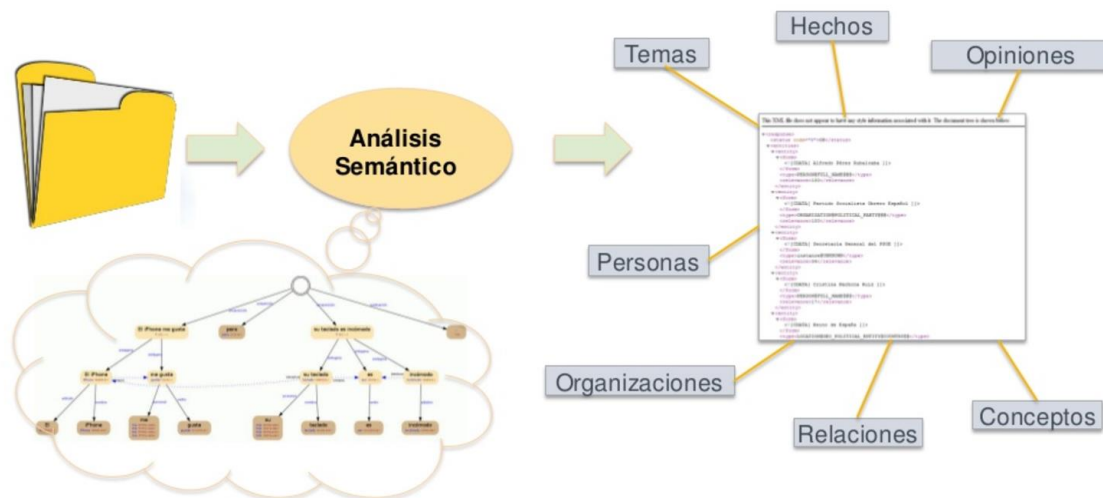


Figura 22. Características extraíbles de las tecnologías MeaningCloud

A continuación, se resumirán los APIs que incluye MeaningCloud y en que campos son útiles cada una de las herramientas para el análisis de datos.

**Topics Extraction** obtiene la información más relevante del texto en etiquetas, sean personas, lugares, organizaciones, etc. [27] Este etiquetado ayuda a estructurar información no estructurada. Destacar que el análisis también detecta URLs, emails y hashtags, proporcionando valor semántico a los tweets.

**Text Classification** asigna una o varias categorías a un texto para hacer más comprensible su estudio mezclando aprendizaje automático y reglas para clasificar textos con exactitud según categorías predefinidas. Esto proporciona una potencia de clasificación y filtrado de textos. La API se vale de modelos predefinidos de clasificación ampliamente aceptados.

**Sentiment Analysis** permite la extracción de información subjetiva de los datos no estructurados mediante lingüística computacional, análisis de textos y procesamiento del lenguaje natural.

**Text Clustering** permite la distribución de un conjunto de textos en varios grupos (Clústeres) dadas las similitudes y diferencias presentes entre ellos nombrando cada uno con un nombre representativo. No requiere taxonomías predefinidas, es decir, emplea el aprendizaje no supervisado para descubrir patrones en los textos y agruparlos basándose en la similitud de contenido entre ellos.

## MODELOS DE CLASIFICACIÓN

La clasificación de textos es un sistema que permite clasificar contenido acorde a unas categorías después de una etapa de entrenamiento. MeaningCloud ha establecido este sistema a través de un modelo de clasificación que se puede observar en la figura 23, siguiendo el siguiente flujo de procesado.

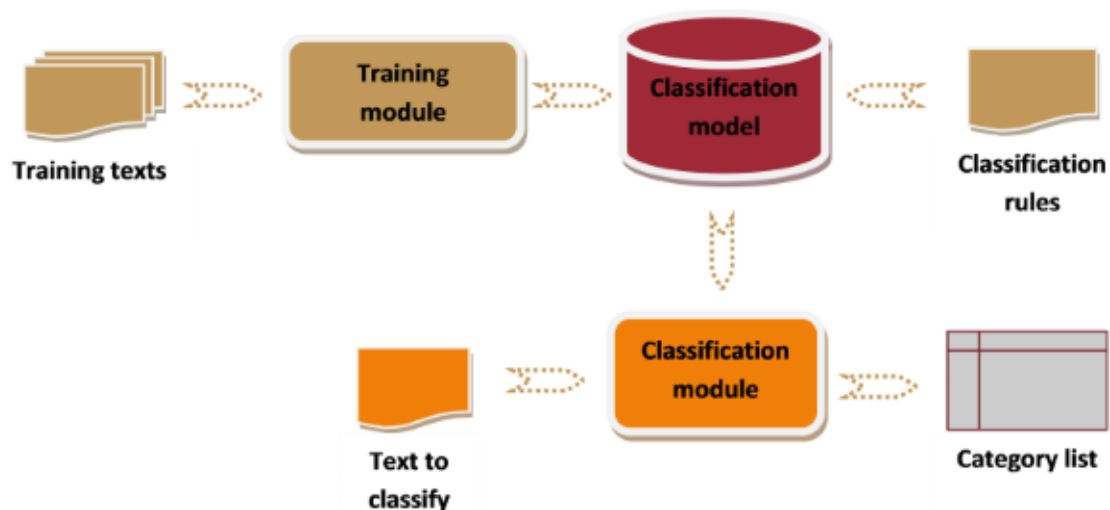


Figura 23. Sistema de clasificación MeaningCloud

Como se puede observar se basa en dos principios. Por una parte, el modelo de entrenamiento obtenido a través de una etapa de entrenamiento. Sumándole por otro lado unas reglas de clasificación que en función de la situación puede utilizar distintas reglas.

Como usuario podemos participar en la etapa de entrenamiento de los datos, en las reglas de clasificación o en ambas. Por lo tanto, tenemos tres tipos de clasificadores.

- **Statistical Model (Modelo estadístico)**

Se compara exclusivamente el texto de entrada con los textos de entrenamiento incluidos en el modelo. Cuando se entrena un modelo, se asocia cada categoría a ejemplos del texto, de modo que, al introducir un texto para ser clasificado, el sistema lo comparará con los ejemplos y determinará cuál es el más cercano.

- **Rule-based Model (Modelo basado en reglas)**  
Emplea exclusivamente reglas de clasificación. Esto implica que la clasificación se realiza considerando solamente términos definidos en el modelo.
- **Hybrid Model (Modelo híbrido)**  
Como su propio nombre indica es una versión híbrida de los dos modelos descritos anteriormente.

Es importante dejar que un modelo de clasificación comprende la lista de categorías, así como los recursos (algoritmos) necesarios para clasificar los documentos en las clases definidas. Este algoritmo combina métodos estadísticos con reglas lingüísticas para obtener la máxima precisión en la clasificación y control sobre los resultados. Adicionalmente contiene cada categoría un conjunto de campos adicionales para suministrar texto de entrenamiento.

De los modelos que proporciona MeaningCloud se ha decidido hacer uso del modelo **Social Media**. Este modelo proporciona una clasificación simple para categorizar de forma comprensiva todos los mensajes de medios sociales que se deseen analizar.

## RESTRICCIONES MEANINGCLOUD

MeaningCloud es un servicio Cloud con diferentes planes de pago y gratuito. Para este proyecto y en concreto las pruebas realizadas, se ha optado por una cuenta gratuita. El plan gratuito que ofrece MeaningCloud presenta restricciones en el tiempo entre peticiones al API y el número total de peticiones por cuenta.

El número total de peticiones posibles para una cuenta gratuita es de 40.000 al mes. Número suficiente para las pruebas realizadas.

Cada segundo sólo permite dos peticiones al API de MeaningCloud. Es otra restricción de cuentas gratuitas. Para evitar errores y retrasos entre peticiones se ha decidido implementar un pequeño método que establece un pequeño margen de milisegundos entre peticiones ajustado para que haga dos peticiones por segundo.

MeaningCloud permite extraer información de datos no estructurados. Una de las principales características por la que se ha elegido MeaningCloud frente a otras librerías ha sido la posibilidad de agrupar el *timeline* del usuario en clústeres dada una similitud entre los tweets. Esto ayuda al resto de análisis para obtener mejores resultados como se verá en el capítulo 5 de esta memoria.

Se han tenido en cuenta otros aspectos como la amplia documentación en diferentes idiomas y la posibilidad de realizar peticiones al API en varios lenguajes de programación.

Otros proyectos como Watson de IBM [29] permiten extraer entidades, conceptos, sentimiento y relaciones entre otras funciones de datos no estructurados, pero las ventajas descritas anteriormente hacen que haya sido MeaningCloud el servicio escogido para este proyecto.

### 4.1.2 ARQUITECTURA

Para el desarrollo de una aplicación web es necesario conocer que la arquitectura en la que están basadas es una arquitectura cliente / servidor, en la que por un lado se tiene al cliente, en este caso el navegador web, y por otro al servidor que es el encargado del procesamiento de los datos. En la figura 24 se presentan los componentes que actúan en la aplicación.

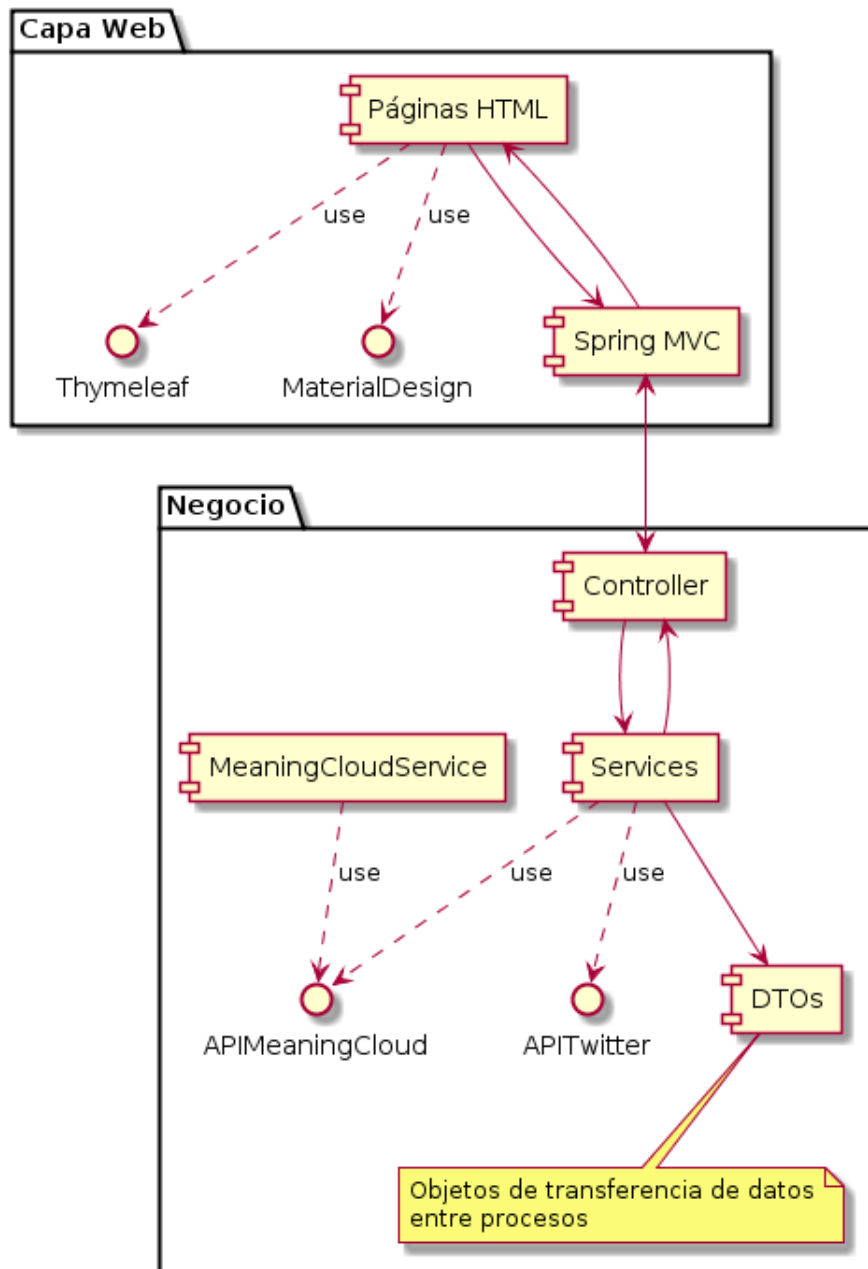


Figura 24. Arquitectura de la aplicación TBotDetector

Los componentes se han dividido en dos paquetes principales. El primero contiene todo lo relacionado con la parte de la vista de la aplicación y el segundo lo forma todo aquello que interacciona con la transformación de los datos.

HTTP es el protocolo cliente / servidor para intercambiar información entre el cliente web (página HTML de la aplicación), y los servidores HTTP. Thymeleaf [22] es

un motor de vistas que permite la visualización de información contenida en objetos enviados desde el controlador. El manejo es sencillo, ahorrando código en las páginas. El uso de Material Design [23] permite diseñar páginas web de forma fácil sin recurrir a extensos archivos CSS. Es importante en el diseño web ya que nos permite adaptar la aplicación a escritorio, Tablet o móvil. El último componente, Spring MVC, permite gestionar las peticiones generadas por la vista para ser redirigidas al controlador correspondiente. En la figura 25 se puede ver como el *Front Controller* es el encargado de redirigir la petición al controlador y la respuesta asociada a la página correspondiente.

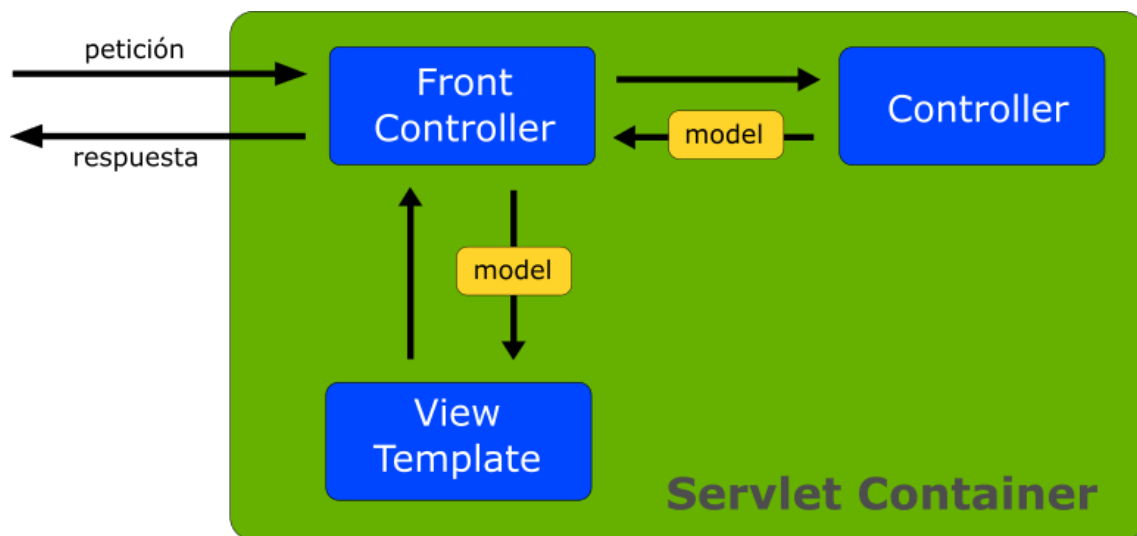


Figura 25. Proceso de petición y respuesta gestionado por Spring MVC

El componente que se encarga del procesamiento de los datos son los servicios de la aplicación. Estos servicios encapsulan en objetos de clases DTO la información que será enviada al controlador. El API de Twitter permite al servicio establecer la conexión para recoger la lista de tweets escritos por los usuarios. Por último, el servicio MeaningCloud es una clase que conecta con el API de MeaningCloud para extraer la información relacionada con los tweets del usuario.

A continuación, se van a detallar las clases que interactúan dentro de los dos componentes principales.

## **CAPA WEB**

se caracteriza por incluir un conjunto de clases encargadas de la configuración del proyecto. Permiten la integración entre Thymeleaf y Spring MVC para que la vista pueda enviar y recibir objetos desde los controladores. La figura 26 muestra el diagrama de clases.

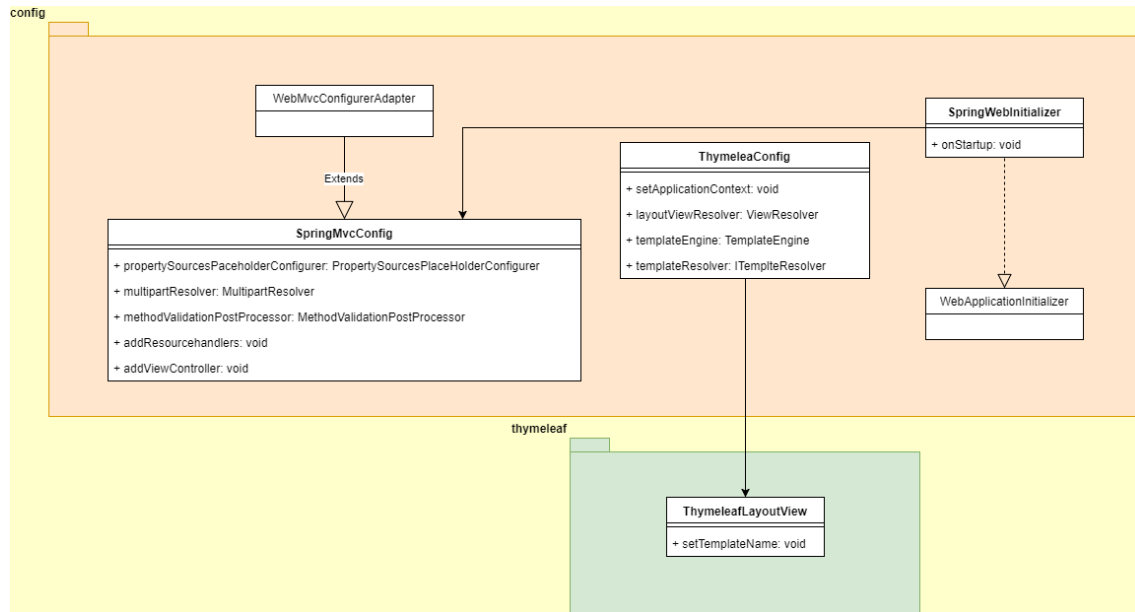


Figura 26. Diagrama de clases encargadas de la configuración de la capa web

La clase **SpringWebInitializer** es la encargada de crear el contexto de aplicación de Spring (Spring application context). Registra **SpringMvcConfig**, es decir, proporciona los Beans [36] necesarios para la configuración y funcionamiento de Spring.

Destacan el Bean para poder enviar archivos al servidor, el Bean que proporciona la ruta donde se encuentran los archivos estáticos de la aplicación (CSS) para almacenarlos en una caché y evitar la carga sucesiva de estos archivos de estilos cada vez que la aplicación se ejecute. Y por último, el controlador al que se debe redirigir la página de inicio de la aplicación.

Otro elemento importante es la creación de un Dispatcher Servlet para el mapeo de vistas controlador.

**ThymeleafConfig** crea la configuración de Thymeleaf. Asigna la ruta donde se encuentran los archivos .html e indica que para esta configuración los archivos de la vista serán de tipo HTML.

Gracias a la clase **ThymeleafViewResolver** es posible incorporar un potente controlador de templates para proporcionar una estructura de vistas común para todos los controladores. Permite establecer un layout principal que permanecerá en todas las vistas de la aplicación y se irán añadiendo fragmentos o vistas a este layout en función de las necesidades de cada página.

La clase **ThymeleafLayoutView** es añadida a través de un objeto **ViewResolver** para establecer el mapeo de vistas explicado anteriormente a través del método `setTemplateName(String templateName)`. Este String es la url que establece donde se encuentran las vistas, qué vista es incluida como layout principal y cuál es la vista que se incorpora dentro de esta principal.

Este layout principal es usado principalmente para mantener la parte web que siempre será visible independientemente de la página en la que se encuentre el usuario. Como por ejemplo el menú y cabecera. Sin embargo, también se añaden archivos de configuración de librerías externas en esta página.

## **CAPA DE NEGOCIO**

Incluye todas las clases encargadas de transformar y modelar los datos que servirán como base del análisis. Para presentar las clases que intervienen en este componente se ha optado por dos diagramas de clases. Corresponden a las clases que intervienen en los dos análisis de la aplicación. Por un lado, aquellas que pertenecen al análisis simple y por el otro al análisis múltiple.

En la figura 27 se pueden apreciar las clases del análisis simple.

La clase **DashboardSimpleController** se encarga de redirigir los datos obtenidos de la vista al servicio principal, al igual que permite el intercambio de datos desde servicio a la capa vista.

Para este intercambio emplea la clase **AnalisisSimpleService**. Es el servicio principal de la aplicación y en él se encuentran los algoritmos que acceden a otros servicios y crean los data transfer (DTO) como contenedores de datos para la vista.

La clase **AnalisisSimpleService** por lo tanto se encarga de obtener los datos asociados a la cuenta del usuario. También emplea la clase **MeaningCloudService** para recoger los datos que proporciona MeaningCloud de cada tecnología (Clústeres, entidades, categorías y análisis de sentimiento) y **CalendarioDTO** para almacenar el estudio de la actividad semanal y diaria del usuario.

Este análisis permite volcar la información de los tweets duplicados en **DuplicateTweetsDTO**.

Y se crean los métodos para sumar toda la información de los análisis y los que realizan la toma de decisiones en base a estos resultados añadiéndose en sus



correspondientes **clases DTO**. Ayuda a **UserTwitterDTO** a dividir la información por características y así tenerla organizada.

Por último, **UserTwitterDTO** almacena un objeto de cada uno de estos análisis. Es la encargada de crear el objeto que contendrá toda la información devuelta de las tomas de decisiones.

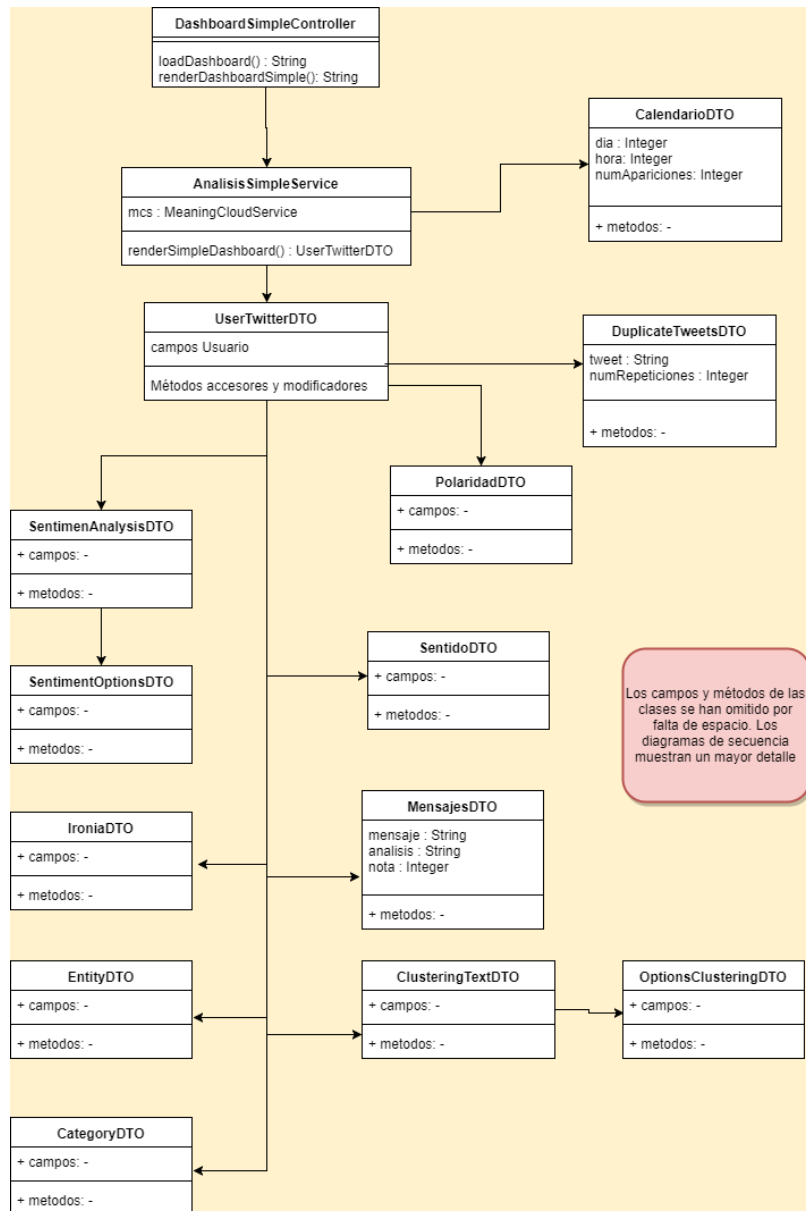


Figura 27. Diagrama de clases para el Análisis Simple

La figura 28 representa las clases que intervienen en la detección de redes de bots.

La clase **DashboardSimpleController** da paso a los datos de la vista a los servicios para procesar la información y también permite la devolución del objeto con la información a la vista. Su método principal `retrieveData()` es el encargado de dar paso a los sucesivos métodos que obtienen y procesan la información.

El servicio con el que se comunica es **AnalisisMultipleService**. Esta clase proporciona los métodos que acceso a la clase **MeaningCloudService** y obtienen la información, en este análisis, de todas las tecnologías salvo el análisis de sentimiento.

Esta información contiene los resultados con los tweets duplicados que son almacenados en una estructura a parte situada en la clase **DuplicateTweetsDTO**

De nuevo, **CalendarioDTO** recoge la información respecto a la actividad semanal y diaria de los usuarios propuestos para el estudio.

Los métodos de recolección de datos y análisis utilizan los **DTO** para almacenar esta información y permiten organizar cada objeto de **UserTwitterRedBotDTO** por características. Ayuda a comprender mejor la comparación entre el usuario bot y el usuario que está siendo estudiado.

**AnalisisMultipleService** se encarga de comparar la información en métodos de la propia clase para obtener los resultados finales a través de este objeto **UserTwitterRedBotDTO**.

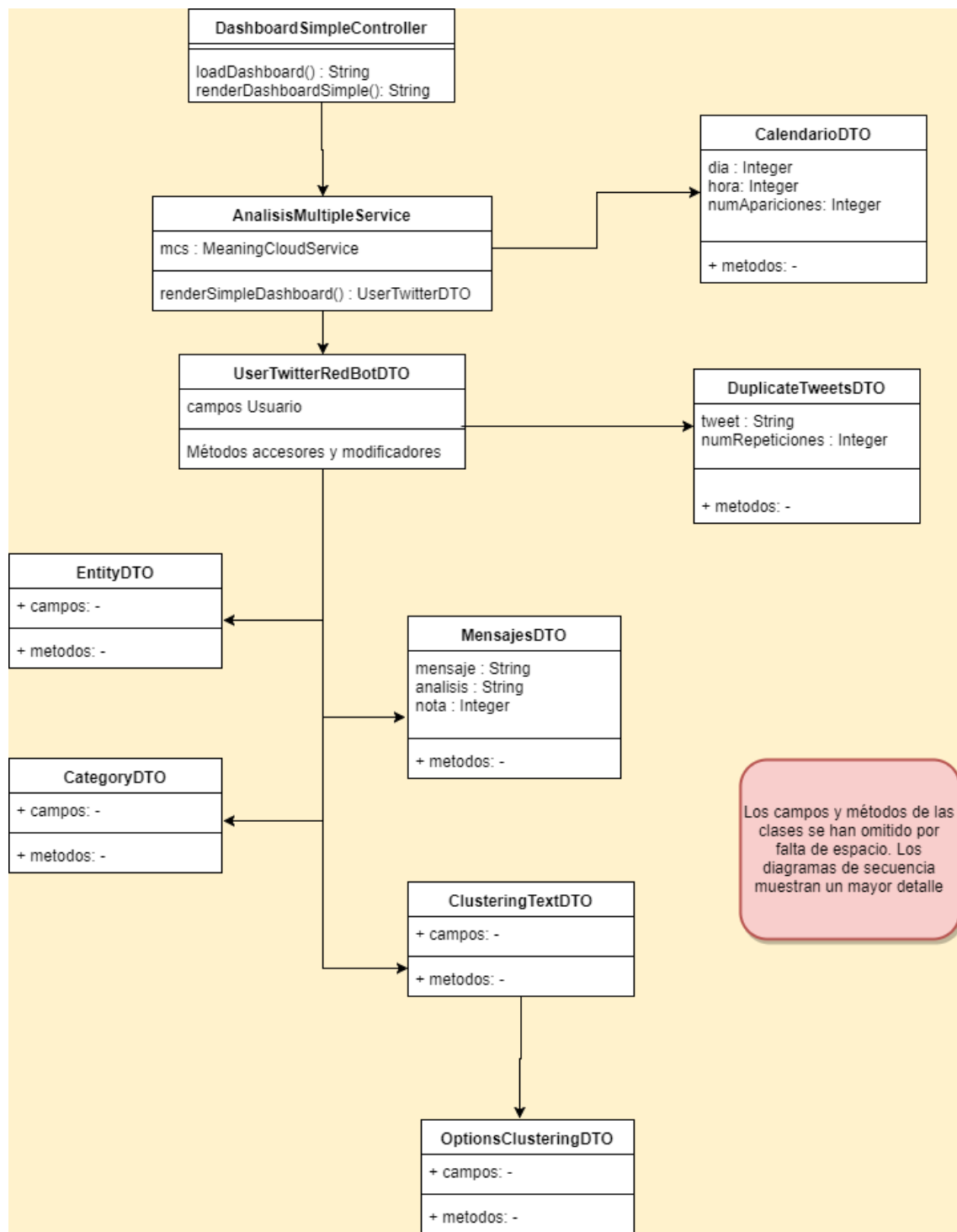


Figura 28. Diagrama de clases para el Análisis Múltiple

### 4.1.3 DIAGRAMAS DE SECUENCIA

Los siguientes diagramas están divididos entre las dos funcionalidades principales de la aplicación. Se describe la interacción de los objetos y en qué orden lo hacen. Para cada análisis, primero se muestra un diagrama de secuencia general del caso de uso y a continuación aquellos diagramas que se creen necesarios para el entendimiento de la funcionalidad. Los diagramas que representan el **análisis simple** para obtener usuarios bot en TBotDetector son los siguientes.

El diagrama de la figura 29 corresponde al caso de uso **Analizar Perfil**. Recoge el proceso que sigue el flujo de la aplicación desde el momento en el que el usuario desea analizar un perfil haciendo clic en el botón de la vista. La interacción es completa entre Vista, Controlador y Modelo. Los datos obtenidos de la vista serán redirigidos al controlador que permite dar paso al servicio que va analizar al usuario. Este servicio, **AnálisisSimpleService**, es el encargado de recuperar la información de la cuenta, comprobar si el usuario debe ser analizado y obtener, si es necesario, el análisis del contenido de los tweets que forman el set de datos del usuario para generar una respuesta a este análisis.

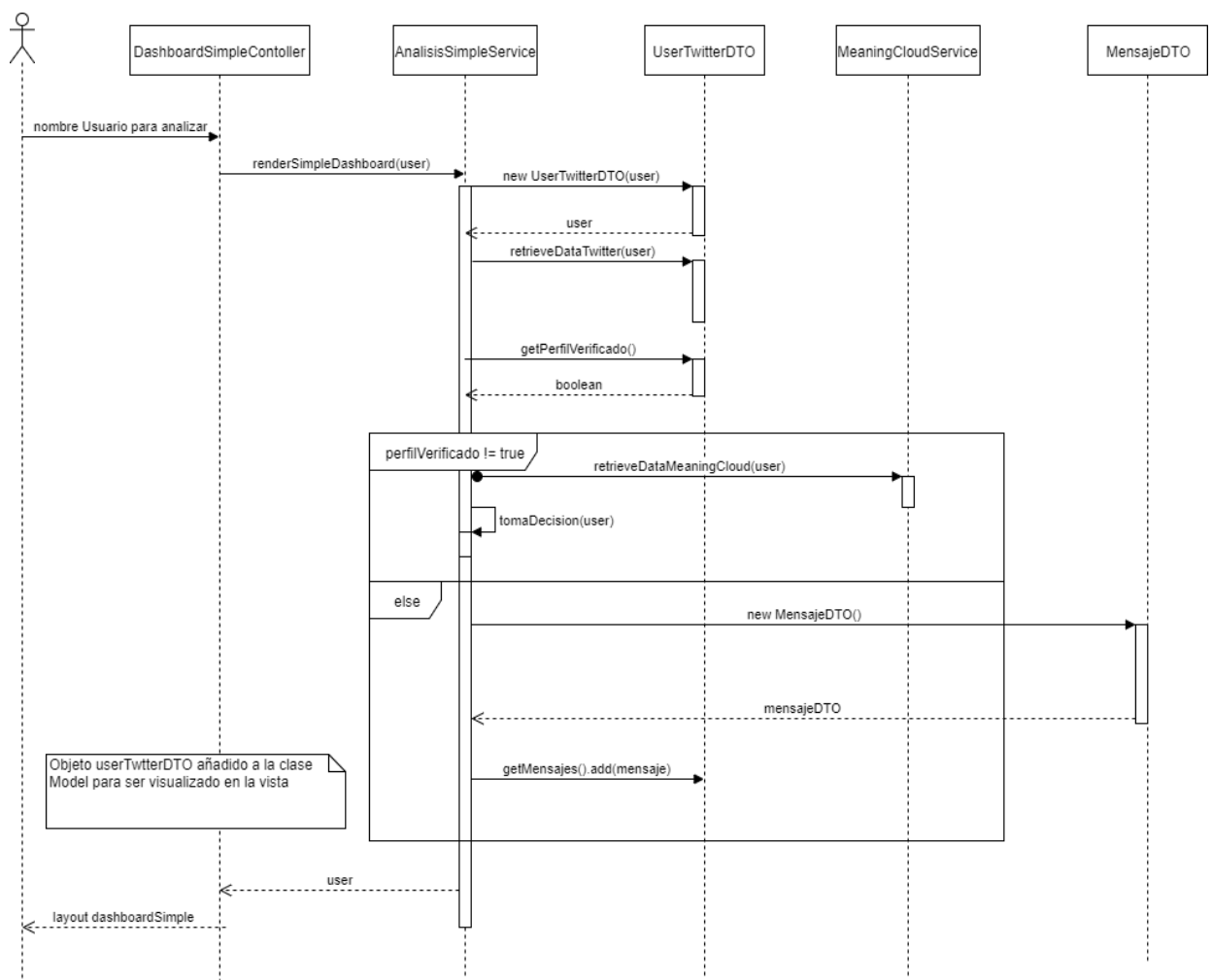


Figura 29. Diagrama de secuencia general para el Análisis Simple

El caso de uso **Recuperar información de la cuenta** está representado por el siguiente diagrama de secuencia. Es el encargado de recoger toda la información de la cuenta del usuario y del conjunto de tweets que se van a analizar. Ver figura 30.

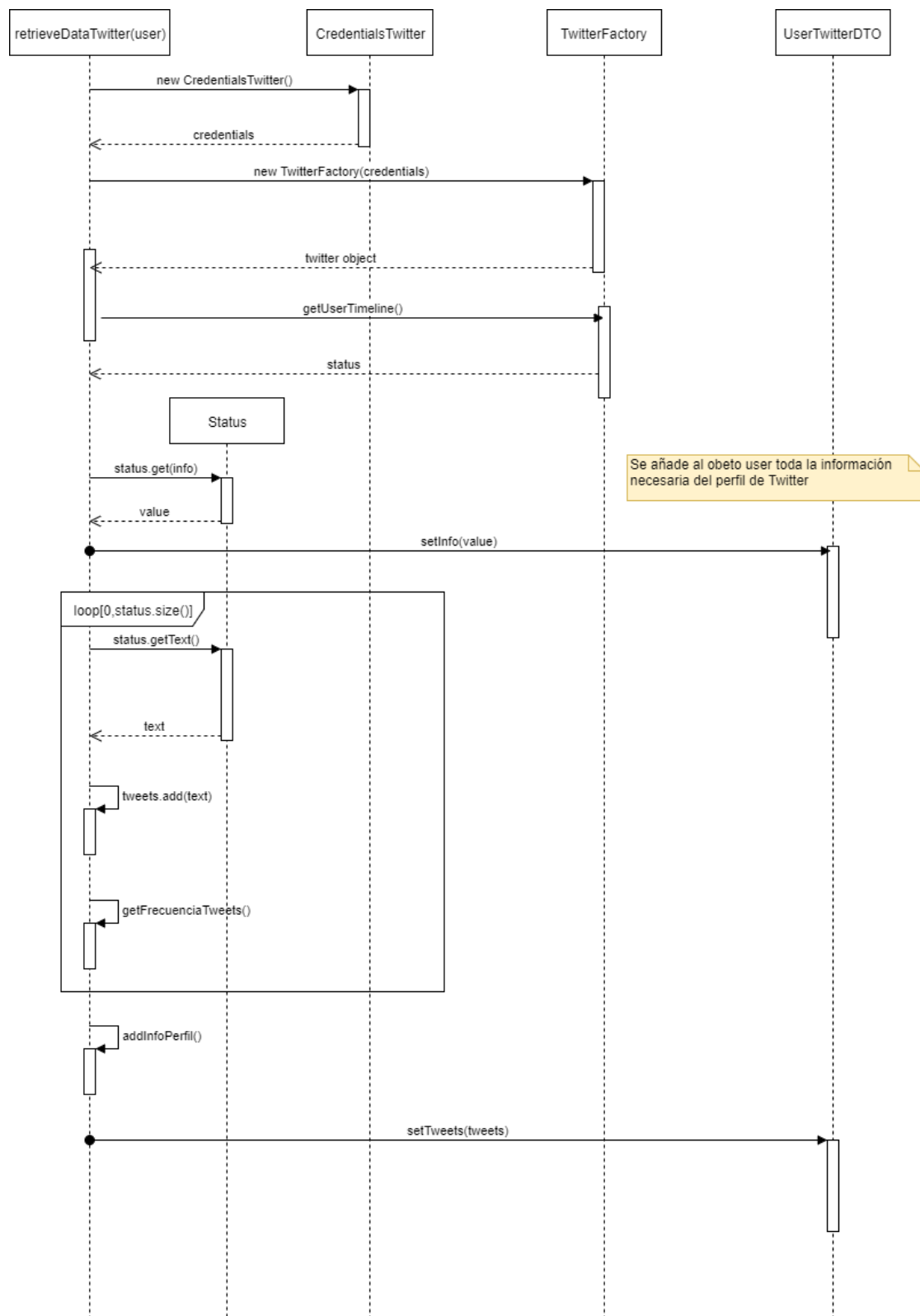


Figura 30. Diagrama de secuencia Recuperar Información de Cuenta

**Procesar información de cuenta** corresponde a la extracción y análisis de los datos obtenidos de MeaningCloud en relación al contenido de la cuenta del usuario. La obtención de clústeres es el primero de los análisis, en función de los resultados obtenidos, se recogen los resultados de la clasificación de textos, entidades y sentimiento de cada uno de los clústeres. Una vez obtenidos, se analizan los resultados de cada clúster para formar un solo conjunto de datos asociado al usuario. La figura 31 representa este proceso.

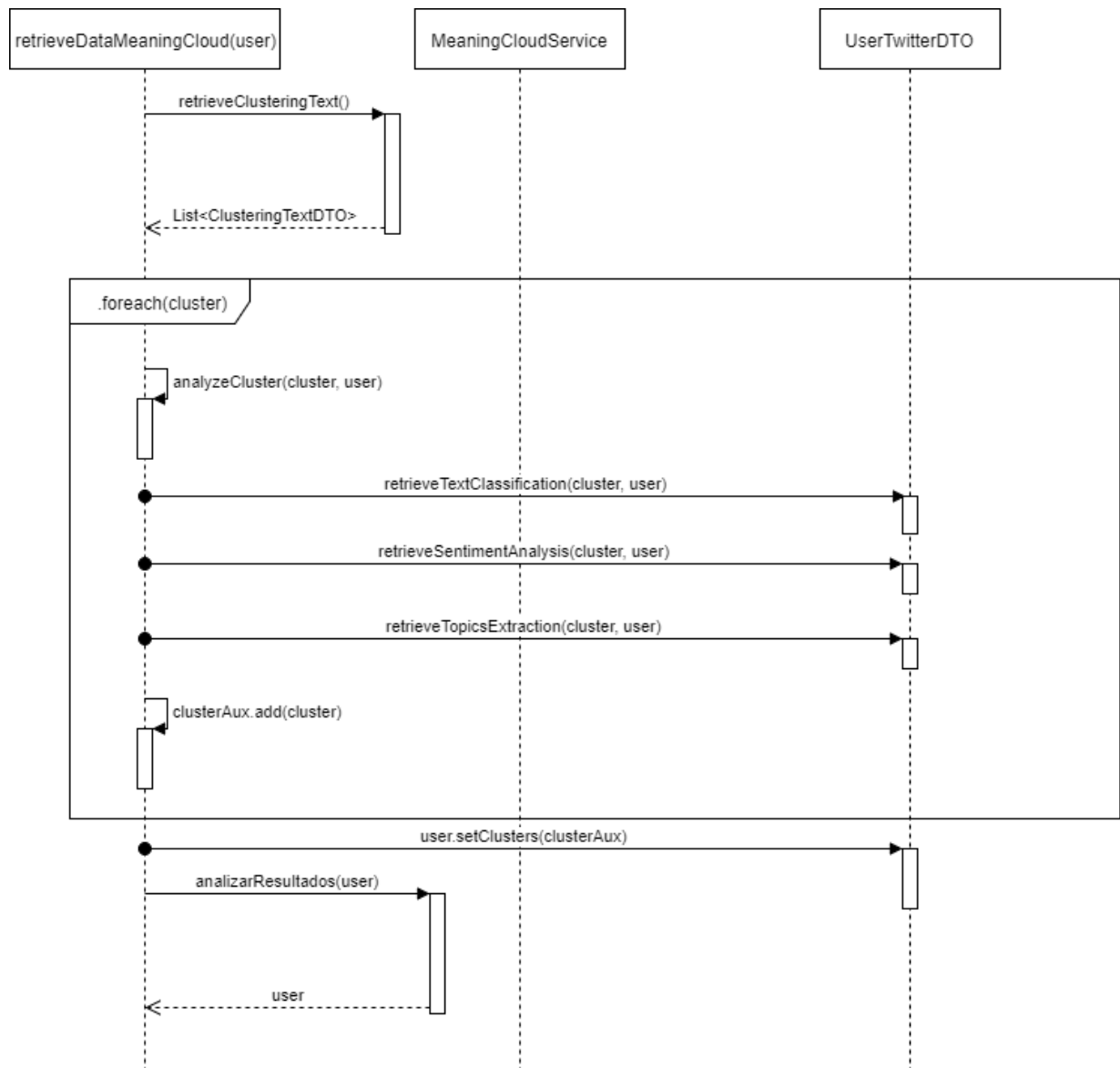


Figura 31. Diagrama de secuencia Procesar Información de Cuenta

Por último, el caso de uso **generar Informe cuenta** corresponde al diagrama de secuencia obtenido de la toma de decisión realizada sobre el usuario. Ver figura 32.

Esta decisión se fundamenta en base a la nota obtenida cada uno de los análisis realizados al usuario. El usuario ya contiene el conjunto de datos asociados a cada clúster agrupados en un mismo objeto **UserTwitterDTO**. Esta unificación de resultados permite tener los datos agrupados para ser más accesibles en la toma de decisión.

El proceso que se encarga de obtener la nota será repetido por cada una de las listas de mensajes de cada uno de los análisis, como se puede apreciar al principio de la figura 32.

Existen mensajes que no proporcionan valor para la toma de decisión, pero sí información para la vista. Estos mensajes serán excluidos para la obtención de la nota final del usuario.

Una vez obtenidas todas las notas, se tiene una lista de objetos **MensajeDTO** con el mensaje asociado al análisis, la conducta del usuario ya sea bot o no y la nota asociada al análisis. Resaltar que una nota elevada significa que la característica tiende a una conducta bot.

Por último, se obtiene mediante comprobaciones simples, la nota del usuario y el veredicto en función de esta. Sin olvidar que el objeto **UserTwitterDTO** sigue conteniendo información valiosa para comprobar que la tendencia del usuario es la correcta en base al resultado final.

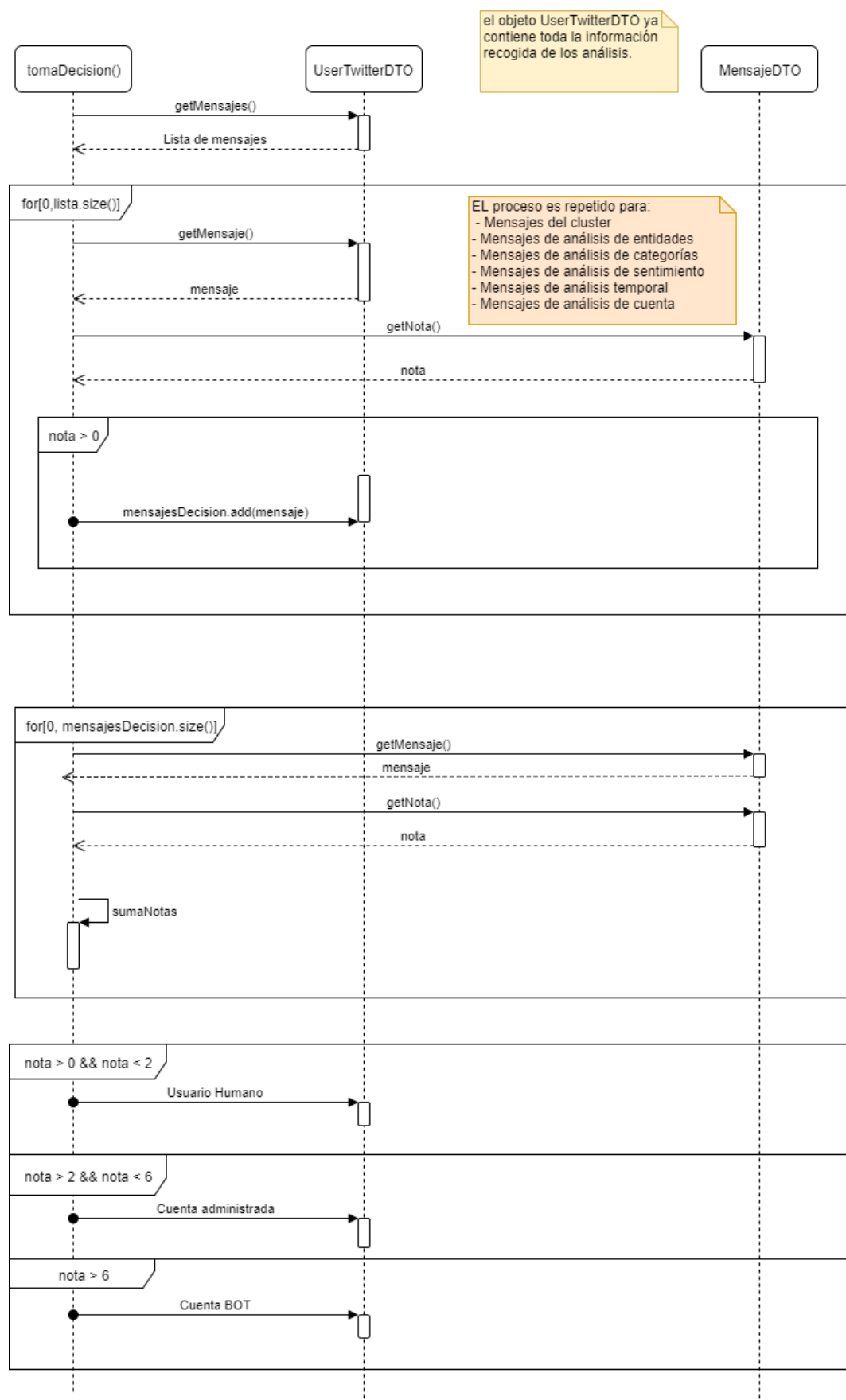


Figura 32. Diagrama de secuencia Generar informe de Cuenta



Los diagramas de secuencia para el análisis de redes de bots en TBotDetector son los citados a continuación.

El primer diagrama de la figura 33 muestra la funcionalidad principal del caso de uso **analizar Red**. Detalla los pasos que se han seguido para detectar si uno o más usuarios pertenece a una red o no. Estos pasos abarcan desde el momento del envío del fichero por parte del usuario de la aplicación, pasando por el controlador encargado del paso de esta información con el servicio correspondiente, en este caso, **AnalisisMultipleService** y la obtención de los análisis requeridos para poder realizar una toma de decisiones lo más eficiente posible. MeaningCloud recogerá los análisis del usuario bot y de los usuarios que sean propuestos para identificar su pertenencia a una posible red de bots.

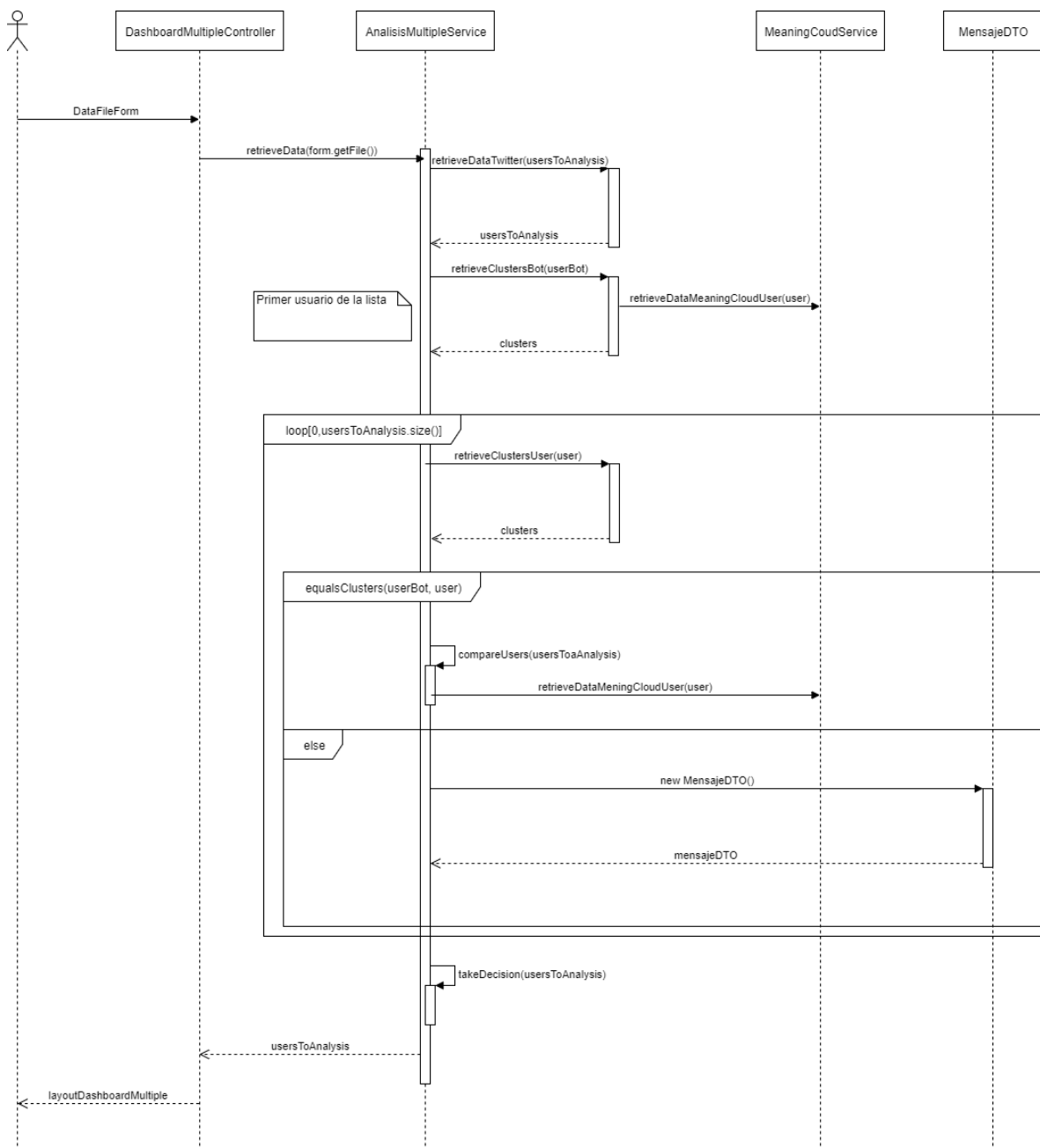


Figura 33. Diagrama de secuencia para el análisis de la red

A continuación, los diagramas de las figuras 34, 35 y 36 representan al caso de uso **recuperar información**. La obtención de los datos de MeaningCloud es común al usuario bot, como al usuario que se desea analizar y comparar. Al igual que ocurría en el análisis simple, primero se obtienen los clústeres más significativos que serán analizados con las tecnologías de extracción de entidades y de categorización de textos. Estos resultados son sumados al objeto **UserTwitterRedBotDTO** para la toma de decisión.

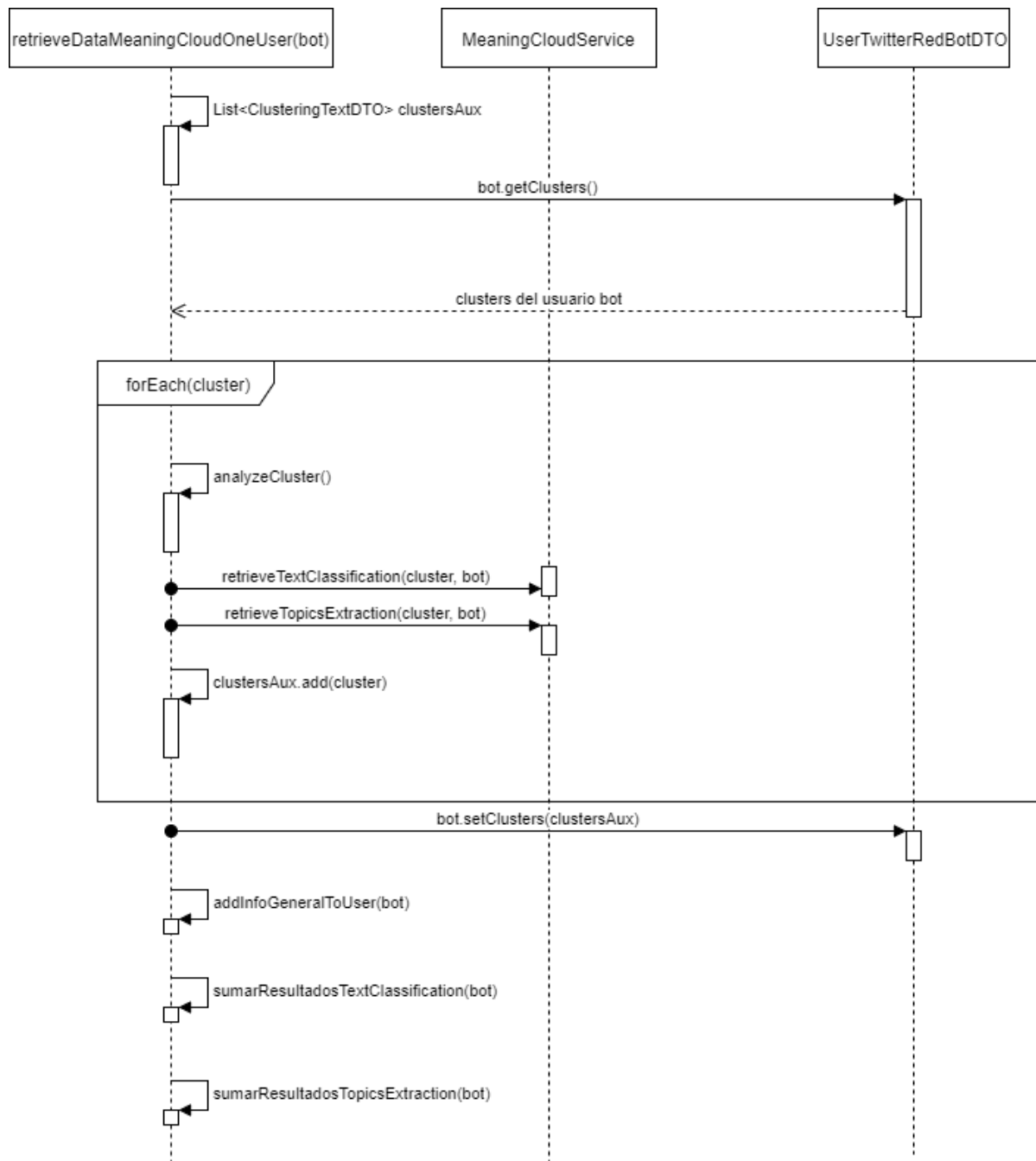


Figura 34. Diagrama de secuencia del caso de uso recuperar información

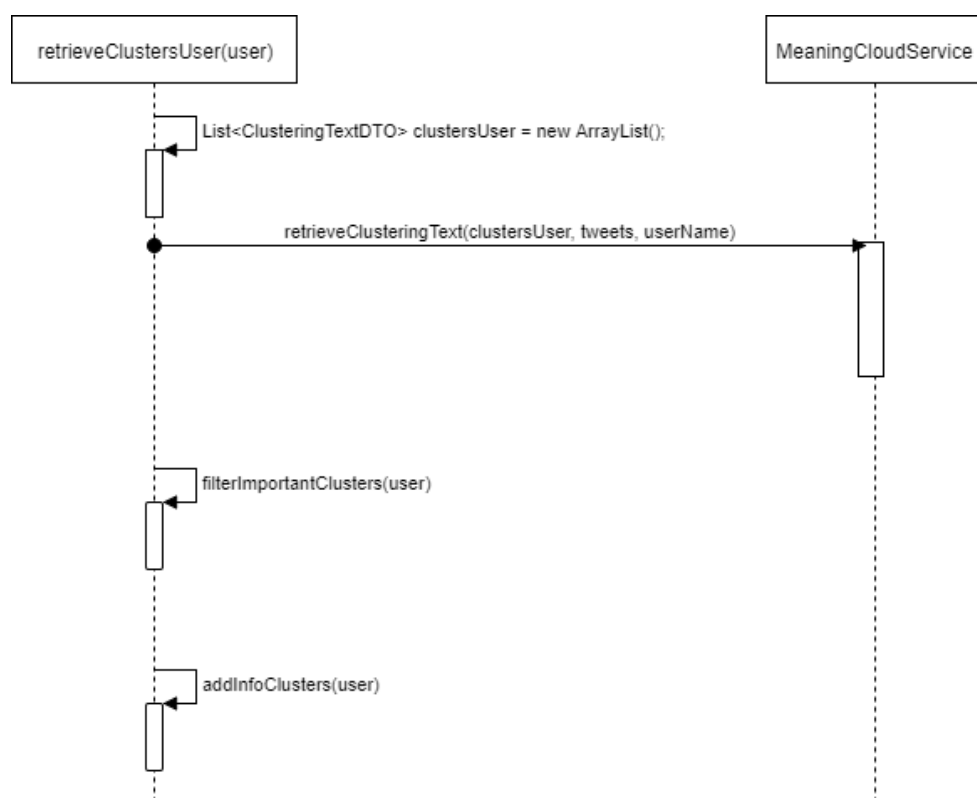


Figura 35. Diagrama de secuencia obtener clústeres usuarios

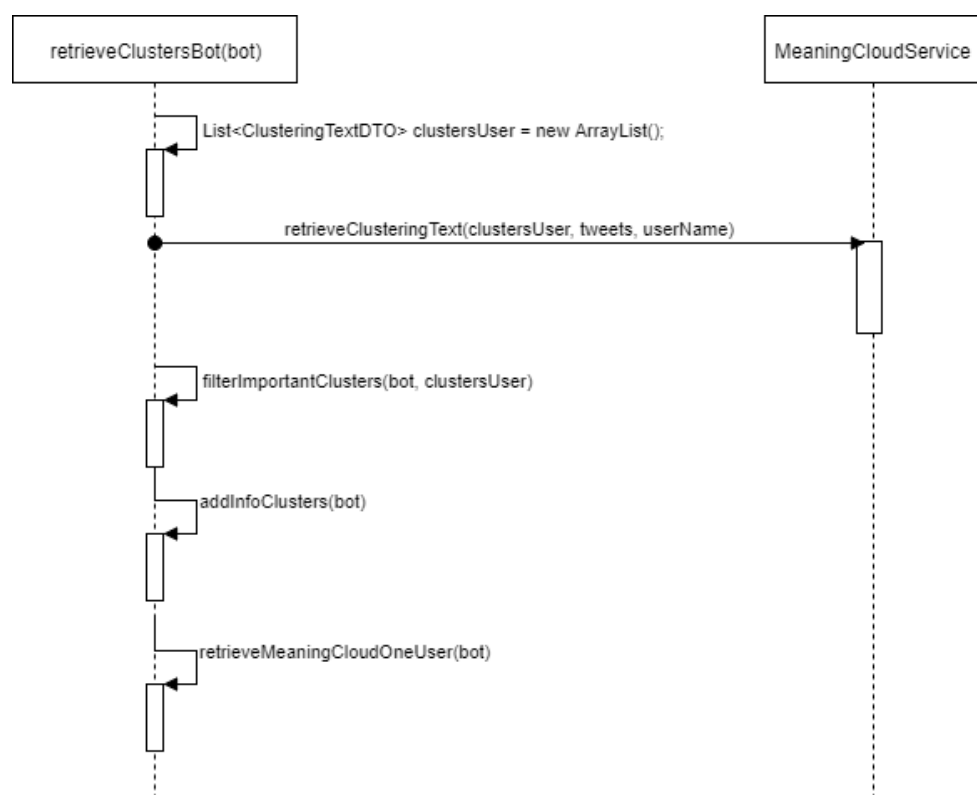


Figura 36. Diagrama de secuencia obtener clústeres bot

Los diagramas 35 y 36 representan la obtención, filtrado y posterior añadido del clúster al usuario y al presunto bot.

El diagrama de la figura 37 representa el caso de uso **procesar información de red**. Permite comparar los resultados entre el usuario bot y el usuario.

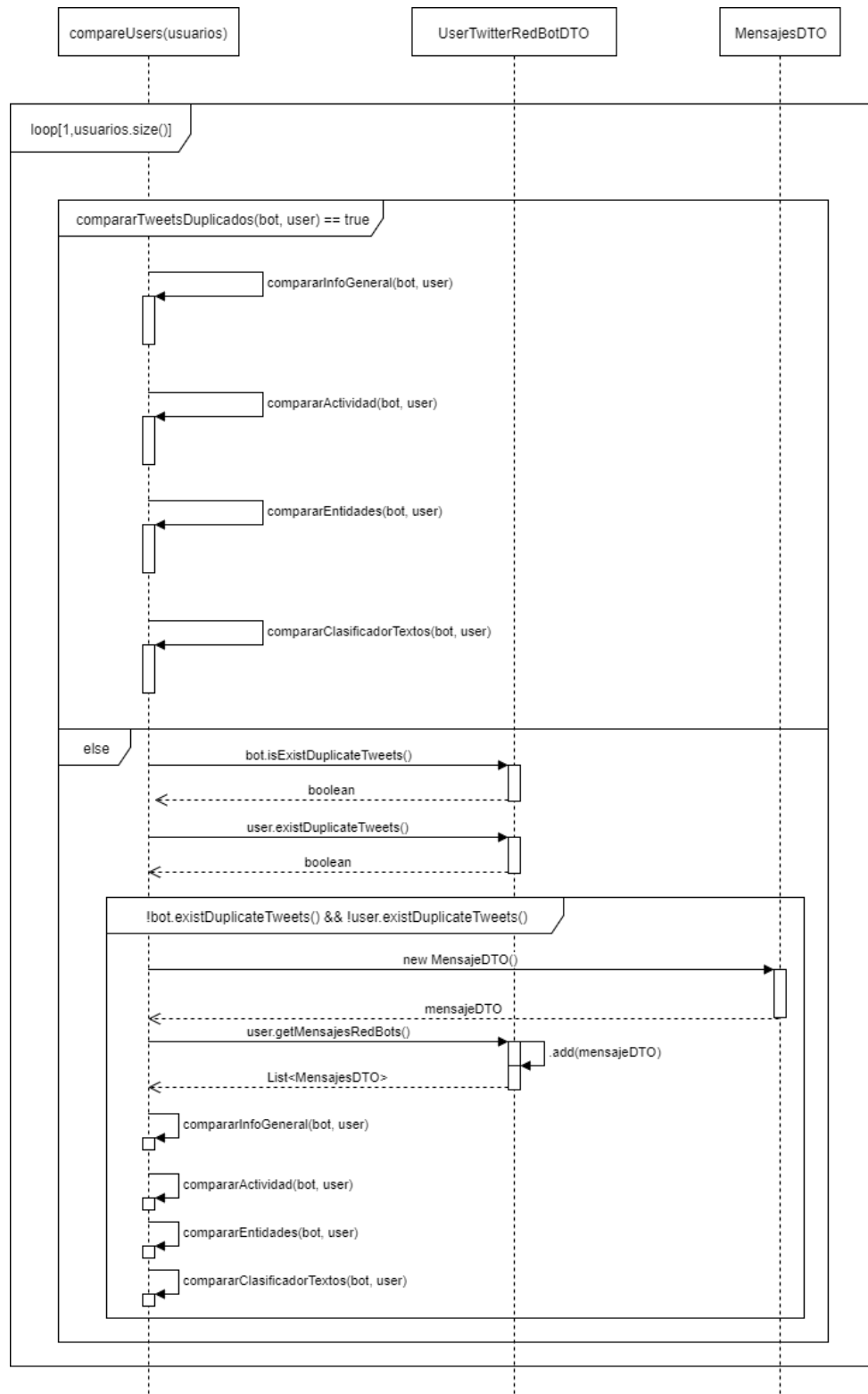


Figura 37. Diagrama de secuencia procesar información de red

Por último, el diagrama de la figura 38 representa el caso de uso **generar informe de red**. Crea una toma de decisión asociada a la comparación de los análisis vistos en la anterior figura. Recoge en primer lugar los mensajes con tendencia a formar parte de la red del usuario bot, para comprobarlos a continuación y poder afirmar que el usuario pertenece o no a la misma red de bots.

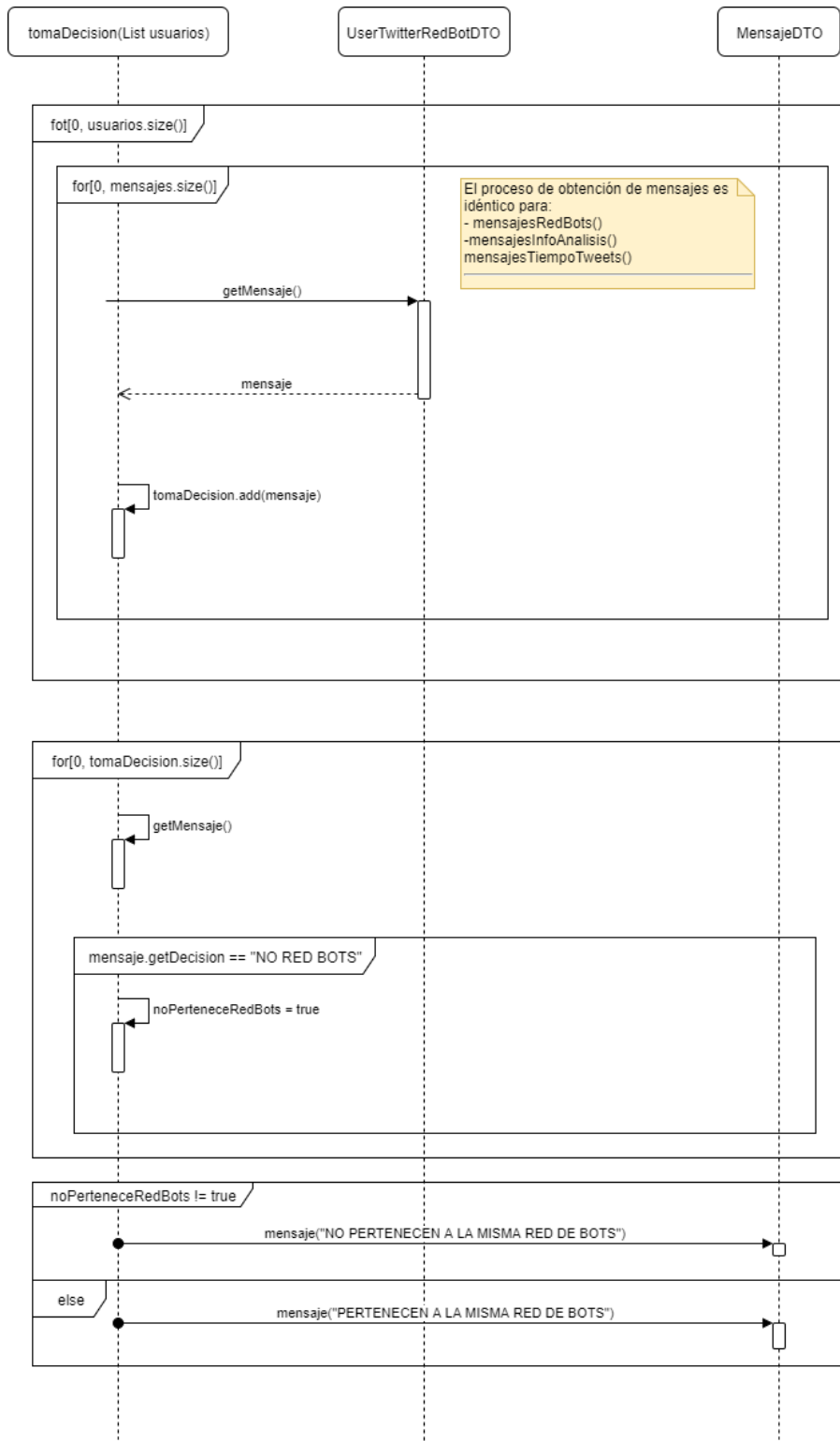


Figura 38. Diagrama de secuencia generar informe red

# Capítulo 5

## EXPERIMENTOS

En este capítulo se detallan los experimentos llevados a cabo con la aplicación para la detección de perfiles bot. Para ello, se ha empleado un conjunto de cuentas de perfiles distintos para averiguar si cumplen o no las características de un usuario bot. Se han formado conjuntos de datos con perfiles reales y que están activos, además de perfiles bot.

Los datos son mostrados a través de la salida de la herramienta directamente tal y como se obtienen por la consola ya que es más sencillo de visualizar actualmente dado el volumen de datos actual, que con el uso de una interfaz gráfica. Se hará especial énfasis en trabajos futuros con esta interfaz gráfica para obtener todo el valor posible de los datos analizados ya que no se ha podido completar a tiempo. Análisis como por ejemplo el relacionado con la actividad del usuario en Twitter, o los porcentajes asociados al análisis del sentimiento del usuario que se verán potenciados gracias a esta funcionalidad, entre otros.

### 5.1 ANÁLISIS SIMPLE

Los experimentos para el análisis simple siguen los pasos marcados en los requisitos de esta memoria.

Cada una de las características arroja una nota que será mayor si es una conducta bot o menor si no lo es. Aquellos mensajes con nota igual a 0 se excluyen del análisis ya que sólo es información para la interfaz.

#### 5.1.1 EXPERIMENTOS CON USUARIOS BOT

##### OBJETIVO

El objetivo es determinar si un usuario que sabemos que es bot es detectado como Bot. Se han realizado experimentos con cuentas bot que administran su cuenta para determinados propósitos en función de las características que presente dicho

bot. Para el análisis se presentan varios usuarios bot que pertenecen a redes bot distintas y es posible diferenciar comportamientos entre ellos.

## DATOS

Se han empleado tres cuentas bot para valorar su estatus. Son, @AbrahamPriego96 y @blogAceiteOliva. Para cada uno de ellos se han descargado los 250 tweets más recientes.

### **HIPÓTESIS BOT 1 - @AbrahamPriego96**

Es un bot de Twitter y se han obtenido los datos de la cuenta y una muestra de 250 tweets para los análisis. A continuación, se detallan los resultados del experimento para determinar su conducta bot y conclusiones en base al análisis.

#### **Análisis realizados – sets de tweets**

El usuario presenta tweets repetidos. Esta característica es importante ya que es una de las características más destacadas de los bots. Ver figura 39.

---

TWEET REPETIDO: ??? | con un total de 10 repeticiones  
TWEET REPETIDO: Hoy en #andaluciadirecto #PriegoDeCordoba #córdoba #HotelRural... <https://t.co/WF08Glr7ht> | con un total de 2 repeticiones

Figura 39. Lista de tweets duplicados del usuario @AbrahamPriego96

#### **Análisis realizados - proporciones**

El análisis indica una alta proporción de menciones y Hashtags y una baja proporción de RT y URL.

La cuenta tiende a resaltar determinados usuarios y Hashtags como veremos en el análisis de entidades, formando menciones a otros usuarios y Hashtags las 5 entidades más importantes. En la figura 40 se observan estos datos.

MENSAJE: Media proporción de RT. El 36.36%. NOTA: 4  
MENSAJE: Alta proporción de Hashtags. El 81.82 %. NOTA: 10  
MENSAJE: Media proporción de URL. El 27.27%. NOTA: 4  
MENSAJE: Alta proporción de menciones. El 100.0%. NOTA: 10

Figura 40. Proporciones del usuario @AbrahamPriego96

## Análisis realizados - entidades

Pasamos al nombrado análisis de entidades. Como se demuestra en la figura 41 existe un alto uso de menciones y Hashtags. Es evidente que el usuario intenta influenciar en determinadas entidades dentro de la red social.

MENSAJE: Existe al menos una entidad cantidad elevada de apariciones. NOTA: 7  
MENSAJE: Una de las entidades mas destacadas @AbrahamPriego96, con un total de 93 apariciones, son menciones a otro usuario. NOTA: 0  
MENSAJE: Una de las entidades mas destacadas, #PriegoDeCordoba con un total de 65 apariciones, es un Hashtag. NOTA: 0  
MENSAJE: Una de las entidades mas destacadas @adirecto, con un total de 24 apariciones, son menciones a otro usuario. NOTA: 0  
MENSAJE: Una de las entidades mas destacadas, #Repost con un total de 15 apariciones, es un Hashtag. NOTA: 0  
MENSAJE: Una de las entidades mas destacadas, #HuertaDeLasPalomas con un total de 10 apariciones, es un Hashtag. NOTA: 0

Figura 41. Análisis de las principales entidades de @AbrahamPriego96

## Análisis realizados - categorías

La poca coherencia en los tweets hace que el resultado para la categorización arroje resultados pobres. Aunque cabe resaltar que el usuario utiliza temas variados en los tweets 42.

MENSAJE: Sólo el 4.37% de los tweets se han podido categorizar como medio ambiente, meteorología y energía. NOTA: 0  
MENSAJE: Sólo el 2.18% de los tweets se han podido categorizar como política. NOTA: 0  
MENSAJE: Sólo el 1.75% de los tweets se han podido categorizar como deporte. NOTA: 0  
MENSAJE: Sólo el 1.75% de los tweets se han podido categorizar como arte y cultura. NOTA: 0  
MENSAJE: Sólo el 1.75% de los tweets se han podido categorizar como economía y finanzas. NOTA: 0  
MENSAJE: Timeline del usuario con diversidad de temas. NOTA: 3  
MENSAJE: El análisis no ha podido clasificar correctamente. Tan sólo un 11.79% de los tweets ha sido posible categorizar. NOTA: 7

Figura 42. Análisis de las principales categorías obtenidas @AbrahamPriego96

## Análisis realizados – actividad

MENSAJE: Frecuencia elevada de tweets en el día Martes, a la hora 18. NOTA: 0  
MENSAJE: Frecuencia elevada de tweets en el día Martes, a la hora 13. NOTA: 0  
MENSAJE: Frecuencia elevada de tweets en el día Martes, a la hora 20. NOTA: 0  
MENSAJE: Frecuencia elevada de tweets en el día Miércoles, a la hora 16. NOTA: 0  
MENSAJE: Frecuencia elevada de tweets en el día Miércoles, a la hora 20. NOTA: 0  
MENSAJE: Frecuencia elevada de tweets en el día Miércoles, a la hora 14. NOTA: 0  
MENSAJE: Frecuencia elevada de tweets en el día Miércoles, a la hora 11. NOTA: 0  
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Martes se escribieron 116 tweets. NOTA: 0  
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Miércoles se escribieron 80 tweets. NOTA: 0  
MENSAJE: Se han encontrado picos de actividad continua en una determinada hora del día. NOTA: 10

Figura 43. Análisis de la frecuencia de actividad @AbrahamPriego96

La actividad del usuario es uno de los análisis más importantes junto a la duplicidad de contenido. En este usuario, existe una frecuencia elevada de tweets



seguidos en un mismo día y en una determinada hora de un día en concreto. Estas dos características arrojan un resultado bot en este análisis. Ver figura 43.

### **Análisis realizados – cuenta del usuario**

El resto de análisis corresponden a características del perfil del usuario. Estas características devuelven un resultado que es sumado al resto de análisis previos.

Resaltar en este análisis la mala reputación de la cuenta dada su reputación próxima al 1. Los valores más altos corresponden a los usuarios seguidos por muchas cuentas pero que siguen a pocos usuarios, como también a los usuarios con muy pocos seguidores, pero con una cantidad considerable de amigos. Para esta cuenta consideramos el primero caso, con 7.760 seguidos y tan sólo 264 seguidores. Ver figura 44.

```
MENSAJE: Mala reputación de la cuenta: 0.97%. NOTA: 10
MENSAJE: El Timeline del usuario no contiene tweets duplicados. NOTA: 3
MENSAJE: Proporción elevada de seguidos. Por cada 1 usuario que sigue, le siguen una media de 28.62. NOTA: 10
MENSAJE: Proporción elevada de RT en los tweets analizados. El 77.0%. NOTA: 10
MENSAJE: El usuario no tiene como foto principal la foto por defecto de Twitter. NOTA: 3
MENSAJE: El usuario posee descripción. NOTA: 3
MENSAJE: Geolocalización activa. NOTA: 3
BOT -> NOTA: 5.9
```

**Figura 44. Análisis de la cuenta @AbrahamPriego96**

### **Conclusión del análisis**

El análisis arroja una nota de 6 lo que implica que una cuenta automatizada o controlada por algún usuario con un determinado propósito. El uso elevado de hashtags es un indicio de usuario bot que intenta realizar spam de determinado contenido a través de una etiqueta. También, el alto número de tweets por día y de menciones a otro usuario son claros indicios de esta conducta.

## **HIPÓTESIS BOT 2 - @blogAceiteOliva**

El siguiente usuario pertenece a una red de bots creada por un medio público [10] con fines propagandísticos.

### **Análisis realizados – sets de tweets**

Al igual que ocurría con la primera cuenta analizada, existen tweets repetidos. Sin embargo, este usuario hace un uso mayor de tweets duplicados, siendo muchos de ellos enlaces a otros tweets de usuarios. Ver figura 45.

```
TWEET REPETIDO: RT @HumorRisas 10 animales que se han colado en fotos que te harán reír https://t.co/FbMbK32DPh https://t.co/YYFJ2aZdLf | con un total de 4 repeticiones
TWEET REPETIDO: RT @OcioUltimate Los 15 grandes descubrimientos para ir a Mercadona https://t.co/1NOg7y6txu https://t.co/00cRKvL5iC | con un total de 4 repeticiones
TWEET REPETIDO: RT @GuapasFamosas Las 10 fotos mas sexys de Scarlett Johansson https://t.co/oRuq2xTGRC https://t.co/scFzj2HxJ4 | con un total de 3 repeticiones
TWEET REPETIDO: RT @Sexologianet Los 10 récords sexuales más locos del mundo https://t.co/VLKENQ0E4g https://t.co/nfocuhLUVC | con un total de 3 repeticiones
TWEET REPETIDO: RT @TeleBLOG 15 fotos polémicas de Sofia (GH16) https://t.co/qyXwShre2g https://t.co/0XiB2HCWGQ | con un total de 3 repeticiones
TWEET REPETIDO: RT @OcioUltimate Los desplantas públicos de Letizia https://t.co/Pt3grBQ0zQ https://t.co/0kzthZ01Jo | con un total de 3 repeticiones
TWEET REPETIDO: RT @BipolarSuelto 10 celebrities que tienen una larga lista de ex https://t.co/A4YsKxJrLK https://t.co/bbXCAFP1BU | con un total de 2 repeticiones
TWEET REPETIDO: RT @OcioUltimate Mentiras que todos decimos en la intimidad https://t.co/0BokuMLTbB https://t.co/Eh11TZxHQI | con un total de 2 repeticiones
TWEET REPETIDO: RT @bipolarsuelto 8 famosas que eran plana y se operaron quedando INFARTANTES https://t.co/ulhFRTkiyJ https://t.co/G9ENya9031 | con un total de 2 repeticiones
TWEET REPETIDO: RT @TeleBLOG Las 10 mejores posiciones para el sexo según 50 Sombras de Grey https://t.co/1aGUUPwbTN https://t.co/YdROXmuI7x | con un total de 2 repeticiones
TWEET REPETIDO: RT @TeleBLOG Quien es el padre del Chavo del 8. https://t.co/2lWZhqdhQe https://t.co/PuUTiWRm48 | con un total de 2 repeticiones
TWEET REPETIDO: RT @GuapasFamosas Scarlett Johansson es la mujer más sexy del mundo https://t.co/oRuq2xTGRC https://t.co/CBtza7MV1 | con un total de 2 repeticiones
TWEET REPETIDO: RT @TeleBLOG Estibaliz Sanz y otros personajillos de la tele que desaparecieron https://t.co/jyQFRHvMqa https://t.co/52iVSvmUJw | con un total de 2 repeticiones
TWEET REPETIDO: RT @GuapasFamosas 10 cosas que hacen a Chloë Moretz la mujer más buscada de Google https://t.co/ZKg0rWCTQK https://t.co/g3ejrfq59d | con un total de 2 repeticiones
TWEET REPETIDO: RT @TeleBLOG Tronistas y pretendientes de 'MYHYV' que trabajaron con Torbe https://t.co/Ewz40FWb2F https://t.co/9D80RwAvR7 | con un total de 2 repeticiones
TWEET REPETIDO: RT @GuapasFamosas 10 famosas españolas con más seguidores en Instagram https://t.co/MF9UJrhyJ2 https://t.co/rthBgqJ5Ca | con un total de 2 repeticiones
TWEET REPETIDO: RT @Sexologianet ¿Qué es el pegging? Conoce esta practica poco común https://t.co/16tkv5nVtt https://t.co/06hVxNnMUT | con un total de 2 repeticiones
TWEET REPETIDO: RT @HumorRisas Los 25 peores tatuajes del mundo. ¡¡¡ Que risa con el número 13!!! https://t.co/tT7qkCrh0A https://t.co/WtLrQpbaFK | con un total de 2 repeticiones
TWEET REPETIDO: RT @GuapasFamosas 10 imágenes que demuestran que Scarlett Johansson gana con los años https://t.co/KP5XEelzzz https://t.co/6G04jEJjqF | con un total de 2 repeticiones
TWEET REPETIDO: RT @Sexologianet Las 10 posiciones sexuales más odiadas por las mujeres https://t.co/vkqxQRj03M https://t.co/3NroNthZ0Q | con un total de 2 repeticiones
TWEET REPETIDO: RT @BipolarSuelto Casa Real: Las fotos que no quieren que descubras https://t.co/Bt3tw0QXB https://t.co/Pm37kMF0fu | con un total de 2 repeticiones
TWEET REPETIDO: RT @OcioUltimate Mira los 10 desplantas mas desubicados de la Reina Letizia https://t.co/Pt3grBQ0zQ https://t.co/0DeAmm2fNU | con un total de 2 repeticiones
TWEET REPETIDO: RT @GuapasFamosas Las 15 grandes meteduras de pata de la Reina Letizia https://t.co/olp3q9kx3t https://t.co/Xun9me3pbf | con un total de 2 repeticiones
```

Figura 45. Lista de tweets repetidos @blogAceiteOliva

### **Análisis realizados – proporciones**

Alta proporción de RT, alta proporción de URL y alta proporción de menciones a otros usuarios. Otro indicio de que la cuenta realiza RT masivos a determinados usuarios y determinados tweets 46.

```
MENSAJE: El timeline del usuario contiene tweets duplicados. NOTA: 10
MENSAJE: Alta proporción de tweets RT. El 81.82 %. NOTA: 10
MENSAJE: Baja proporción de Hashtags. El 0.0%. NOTA: 2
MENSAJE: Alta proporción de URL en tweets. El 100.0 %. NOTA: 10
MENSAJE: Alta proporción de menciones. El 90.91%. NOTA: 10
```

Figura 46. Proporciones del usuario de @blogAceiteOliva

## Análisis realizados – entidades

Las entidades más destacadas son todas menciones a otros usuarios salvo una URL. Se confirman los porcentajes obtenidos anteriormente y la conducta bot del usuario 47.

MENSAJE: Una de las entidades mas destacadas @GuapasFamosas, con un total de 69 apariciones, son menciones a otro usuario. NOTA: 0  
MENSAJE: Una de las entidades mas destacadas @OcioUltimate, con un total de 39 apariciones, son menciones a otro usuario. NOTA: 0  
MENSAJE: Una de las entidades mas destacadas @TeleBLOG, con un total de 37 apariciones, son menciones a otro usuario. NOTA: 0  
MENSAJE: Una de las entidades mas destacadas @BipolarSuelto, con un total de 26 apariciones, son menciones a otro usuario. NOTA: 0  
MENSAJE: Una de las entidades mas destacadas @HumorRisas, con un total de 17 apariciones, son menciones a otro usuario. NOTA: 0  
MENSAJE: Una de las entidades mas destacadas @Sexologianet, con un total de 16 apariciones, son menciones a otro usuario. NOTA: 0  
MENSAJE: Una de las entidades mas destacadas, <https://t.co/Pt3grBQ0zQ> con un total de 11 apariciones, es una URL. NOTA: 0

Figura 47. Análisis de las principales entidades de @blogAceiteOliva

## Análisis realizados – categorías

Poca categorización de los tweets. El alto uso de URL dificulta este análisis, tan sólo categorizando el 8,6 % de un total de 250 tweets analizados. Ver figura 48

MENSAJE: Sólo el 3.17% de los tweets se han podido categorizar como asuntos sociales. NOTA: 0  
MENSAJE: Sólo el 2.26% de los tweets se han podido categorizar como salud. NOTA: 0  
MENSAJE: Sólo el 1.36% de los tweets se han podido categorizar como ocio, estilo de vida y tiempo libre. NOTA: 0  
MENSAJE: Sólo el 0.9% de los tweets se han podido categorizar como religión y credos. NOTA: 0  
MENSAJE: Sólo el 0.9% de los tweets se han podido categorizar como economía y finanzas. NOTA: 0  
MENSAJE: Timeline del usuario con diversidad de temas. NOTA: 3  
MENSAJE: El análisis no ha podido clasificar correctamente. Tan sólo un 8.6% de los tweets ha sido posible categorizar. NOTA: 7

Figura 48. Análisis de las principales categorías de @blogAceiteOliva

## Análisis realizados – actividad

Una frecuencia de tweets continúa y siguiendo una tendencia como se puede apreciar en la siguiente figura 49. Estas semejanzas en los resultados implican una monitorización en la cuenta.

MENSAJE: Frecuencia elevada de tweets en un mismo día. El Miércoles se escribieron 18 tweets. NOTA: 0  
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Jueves se escribieron 17 tweets. NOTA: 0  
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Viernes se escribieron 18 tweets. NOTA: 0  
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Sábado se escribieron 26 tweets. NOTA: 0  
MENSAJE: Se han encontrado picos de actividad continua en una determinada hora del día. NOTA: 10

Figura 49. Análisis de la actividad de @blogAceiteOliva

## Análisis realizados – cuenta del usuario

Por último, los resultados del análisis del perfil del usuario 50. Presenta una mala reputación de la cuenta y una elevada proporción de usuarios seguidos.

MENSAJE: Mala reputación de la cuenta: 0.78%. NOTA: 10  
 MENSAJE: El Timeline del usuario contiene tweets duplicados. NOTA: 10  
 MENSAJE: Proporción elevada de seguidos. Por cada 1 usuario que sigue, le siguen una media de 3.55. NOTA: 10  
 MENSAJE: Proporción elevada de RT en los tweets analizados. El 98.0%. NOTA: 10  
 MENSAJE: El usuario no tiene como foto principal la foto por defecto de Twitter. NOTA: 3  
 MENSAJE: El usuario posee descripción. NOTA: 3  
 MENSAJE: Geolocalización no activa. NOTA: 6  
 BOT -> NOTA: 7.82

Figura 50. Análisis de la cuenta de @blogAceiteOliva

## Conclusión del análisis

El análisis arroja un resultado elevado que señalaría la posibilidad de estar ante un perfil bot. El perfil tiene una clara intencionalidad de contactar otras cuentas a través de RT a estas cuentas y de duplicado de contenido a base de RT también. Las entidades tienen un alto número de apariciones de otras cuentas, reflejo de elevado porcentaje de RT realizados. Característica que resalta y permite deducir que estos usuarios podrían ser bots y pertenecer a una misma red de bots.

## HIPÓTESIS BOT 3 - @afrasecool

El usuario analizado comparte determinado contenido de otros usuarios que pertenecen a la misma red.

## Análisis realizados – sets de tweets

El análisis de tweets duplicados obtiene numerosos tweets que han sido utilizados más de una vez en la cuenta. La figura 51 sólo recoge una pequeña muestra de los recogidos en el análisis. Destaca la duplicidad de tweets RT de usuarios como @BipolarSuelto, @SuperSarcasmo o @iphone\_noticias. Son indicadores de un comportamiento bot.

TWEET REPETIDO: RT @TeleBLOG: Los 10 momentos ÉPICOS de la televisión en España <https://t.co/G5vIXM5pIW> <https://t.co/x0kgVmVYY> | con un total de 2 repeticiones  
 TWEET REPETIDO: RT @iphone\_noticias 7 impresionantes trucos para tu iPhone <https://t.co/BGni5ubmIV> <https://t.co/RBrLBpP8Sz> | con un total de 2 repeticiones  
 TWEET REPETIDO: RT @Sexologianet Fantasías sexuales más típicas de las mujeres <https://t.co/eXcNBdZzue> <https://t.co/bo811Ij1Ty> | con un total de 2 repeticiones  
 TWEET REPETIDO: RT @Supersarcasmo: Los 10 mejores escotes de Victoria Bárbara <https://t.co/TjVyCLPtjP> <https://t.co/5JvhqdzWo> | con un total de 2 repeticiones

Figura 51. Breve muestra del numeroso contenido duplicado por el usuario

## Análisis realizados – proporciones

El análisis destaca por obtener una alta proporción de menciones dentro de todos los tweets analizados. Si observamos el análisis anterior, guarda relación que se obtenga un alto porcentaje de RT en menciones si se intenta promocionar un determinado medio o usuario. Ver figura 52.

```

MENSAJE: Media proporción de RT. El 160.0%. NOTA: 4
MENSAJE: Baja proporción de Hashtags. El 0.0%. NOTA: 2
MENSAJE: Alta proporción de URL en tweets. El 100.0 %. NOTA: 10
MENSAJE: Alta proporción de menciones. El 70.0%. NOTA: 10

```

Figura 52. Análisis de las proporciones obtenidas de todos los tweets extraídos del usuario

## Análisis realizados – entidades

Las tres entidades más representativas en el análisis, ver figura 53, son menciones a otros usuarios. Estas menciones destacan por corresponder a los tweets que el usuario ha duplicado. Siendo @BipolarSuelto la entidad más nombrada y RT del usuario. Es otro indicio el promover el contenido del usuario mencionado.

```

MENSAJE: Existe al menos una entidad cantidad elevada de apariciones. NOTA: 7
MENSAJE: Una de las entidades mas destacadas @BipolarSuelto, con un total de 91 apariciones, son menciones a otro usuario. NOTA: 0
MENSAJE: Una de las entidades mas destacadas @Sexologianet, con un total de 26 apariciones, son menciones a otro usuario. NOTA: 0
MENSAJE: Una de las entidades mas destacadas @iphone_noticias, con un total de 23 apariciones, son menciones a otro usuario. NOTA: 0
MENSAJE: Una de las entidades mas destacadas, https://t.co/Q5eROBwoUn con un total de 10 apariciones, es una URL. NOTA: 0
MENSAJE: Una de las entidades mas destacadas, https://t.co/lXccz1b4hP con un total de 10 apariciones, es una URL. NOTA: 0

```

Figura 53. Entidades más destacadas del análisis del usuario. Fuerte implicación de otros usuarios en las publicaciones

## Análisis realizados – categorías

El usuario contiene tweets de que engloban una variedad de temas, pero contrasta la poca categorización del contenido. Esto puede ser debido, si nos fijamos en la figura 53 del análisis de entidades, a que dos de las principales entidades más destacadas son enlaces a otras páginas. Esto dificulta la obtención de categorías dentro del análisis y deriva en una conducta bot. Ver figura 54.

```

MENSAJE: Sólo el 3.57% de los tweets se han podido categorizar como ciencia y tecnología. NOTA: 0
MENSAJE: Sólo el 2.55% de los tweets se han podido categorizar como arte y cultura. NOTA: 0
MENSAJE: Sólo el 1.53% de los tweets se han podido categorizar como turismo, viajes y desplazamientos. NOTA: 0
MENSAJE: Sólo el 0.51% de los tweets se han podido categorizar como deporte. NOTA: 0
MENSAJE: Sólo el 0.51% de los tweets se han podido categorizar como policía y justicia. NOTA: 0
MENSAJE: Timeline del usuario con diversidad de temas. NOTA: 3
MENSAJE: El análisis no ha podido clasificar correctamente. Tan sólo un 8.67% de los tweets ha sido posible categorizar. NOTA: 7

```

Figura 54. Escasa categorización de los tweets del usuario debido al alto número de enlaces compartidos

## Análisis realizados – sentimiento

El único análisis positivo del usuario ha sido el análisis de sentimiento representado en la figura 55. Los resultados no concuerdan con el resto de análisis.

MENSAJE: Se realiza un análisis de sentimiento coherente. No existe un alto porcentaje de análisis nulo.. NOTA: 4

Figura 55. Los tweets han podido ser analizados sin obtener un alto porcentaje de nulos

## Análisis realizados – actividad

La actividad relacionada con la publicación de tweets observada en la figura 56 del usuario, arroja resultados evidentes de que el usuario interacciona en unos determinados días del día, posicionando el contenido en las horas del día que mayor impacto produce en el resto de usuarios.

MENSAJE: Frecuencia elevada de tweets en un mismo día. El Miércoles se escribieron 16 tweets. NOTA: 0  
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Jueves se escribieron 17 tweets. NOTA: 0  
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Viernes se escribieron 14 tweets. NOTA: 0  
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Sábado se escribieron 15 tweets. NOTA: 0  
MENSAJE: Se han encontrado picos de actividad continua en una determinada hora del día. NOTA: 10

Figura 56. Análisis de la actividad del usuario con altos picos de actividad en determinados días y horas

## Análisis realizados – cuenta del usuario

Las características de la cuenta del usuario revelan una mala reputación, así como una desnivelada proporción de seguidores. Estos análisis permiten afianzar los análisis anteriormente obtenidos. Ver figura 57.

MENSAJE: Mala reputación de la cuenta: 0.55%. NOTA: 10  
MENSAJE: El Timeline del usuario contiene tweets duplicados. NOTA: 10  
MENSAJE: Alta proporción de seguidos. El 1.2159841479524438%. NOTA: 3  
MENSAJE: Proporción baja de RT en los tweets analizados. El 0%. NOTA: 3  
MENSAJE: El usuario no tiene como foto principal la foto por defecto de Twitter. NOTA: 3  
MENSAJE: El usuario posee descripción. NOTA: 3  
MENSAJE: Geolocalización no activa. NOTA: 6  
BOT -> NOTA: 6.0

Figura 57. Análisis de la cuenta del usuario. Se obtienen diferentes características para definir la conducta del usuario asociada a la cuenta

## Conclusión del análisis

Los informes señalan que este usuario podría ser un bot. Tiene una resaltada intencionalidad dentro de la red social, posicionando al usuario @BipolarSuelto en determinadas franjas horarias. También se observa como el usuario intenta expandir el contenido a base de RT a otros usuarios que pertenecen a la misma red del usuario, llegando a compartir tweets repetidos en los que estos usuarios son RT masivamente.

### 5.1.2 EXPERIMENTOS CON USUARIOS NO BOT

#### OBJETIVO

Se han realizado experimentos con cuentas de seres humanos que no administran su cuenta para un determinado propósito, si no que utilizan dicha cuenta como personal para interactuar con el resto de usuarios.

#### DATOS

Los usuarios son @SergioRG92, @Caiquet11 y @traedRuffles y se han recogido un set de los 250 tweets más recientes de cada uno.

#### **HIPÓTESIS HUMANO 1 - @SergioRG92**

Es un usuario activo de la red social Twitter, dedicado a la interacción con el resto de usuarios tratando diversos temas de la actualidad y de entretenimiento. Ha presentado los siguientes resultados en el análisis realizado por TBotDetector.

#### **Análisis realizados – sets de tweets**

La primera característica analizada del usuario revela que los sets de tweets analizados están libres de duplicados, ver figura 58. Por lo tanto, es el primer indicio de que la cuenta puede no ser un bot.

**MENSAJE:** El timeline del usuario no contiene tweets duplicados. NOTA: 3

Figura 58. Análisis de tweets duplicados del @SergioRG92 sin muestra de tweets duplicados

### **Análisis realizados – proporciones**

MENSAJE: Alta proporción de tweets RT. El 80.0 %. NOTA: 10  
MENSAJE: Baja proporción de Hashtags. El 0.0%. NOTA: 2  
MENSAJE: Baja proporción de URL. El 10.0%. NOTA: 2  
MENSAJE: Alta proporción de menciones. El 100.0%. NOTA: 10

Figura 59. Proporciones obtenidas de @SergioRG92

El siguiente análisis muestra diferentes proporciones, siendo significativas el RT a otros usuarios y las menciones. Cabe destacar la poca utilización de URL y de Hashtags, es decir, no intenta posicionar determinadas etiquetas o realizar spam a determinados enlaces. Figura 59.

### **Análisis realizados – entidades**

Unido a la alta proporción de menciones, las principales entidades son otros usuarios de Twitter, aunque no existe una evidencia clara exceso de mención en los tweets de algunas de las entidades. Figura 60.

MENSAJE: Una de las entidades mas destacadas @Miotroyo2parte, con un total de 15 apariciones, son menciones a otro usuario. NOTA: 0  
MENSAJE: Una de las entidades mas destacadas @Lassana\_, con un total de 10 apariciones, son menciones a otro usuario. NOTA: 0  
MENSAJE: Una de las entidades mas destacadas @InformaticoFFUU, con un total de 8 apariciones, son menciones a otro usuario. NOTA: 0  
MENSAJE: Una de las entidades mas destacadas @MotivacionesF, con un total de 5 apariciones, son menciones a otro usuario. NOTA: 0  
MENSAJE: Una de las entidades mas destacadas @Sr\_Dios, con un total de 5 apariciones, son menciones a otro usuario. NOTA: 0  
MENSAJE: No existe una entidad representativa dentro de los tweets analizados. NOTA: 3

Figura 60. Análisis de las principales entidades de @SergioRG92

### **Análisis realizados – categorías**

El 22% es un resultado no excesivamente bajo para este análisis, pero se debe considerar insuficiente, aunque cuenta con diversidad de temas dentro de este porcentaje. Figura 61.

MENSAJE: El 21.33% de los tweets se han podido categorizar como deporte. NOTA: 0  
MENSAJE: Sólo el 5.33% de los tweets se han podido categorizar como policía y justicia. NOTA: 0  
MENSAJE: Sólo el 4.0% de los tweets se han podido categorizar como disturbios, conflictos y guerra. NOTA: 0  
MENSAJE: Sólo el 2.67% de los tweets se han podido categorizar como trabajo. NOTA: 0  
MENSAJE: Sólo el 2.67% de los tweets se han podido categorizar como religión y credos. NOTA: 0  
MENSAJE: Timeline del usuario con diversidad de temas. NOTA: 3

Figura 61. Análisis de las principales categorías de @SergioRG92

### **Análisis realizados – sentimiento**

El 37% de los tweets han sido etiquetados como nulos en el análisis de sentimientos. Se encuentra a un 7% de los límites del análisis entre conducta bot, pero en este caso por poco porcentaje el análisis concluye que es una conducta bot. Ver figura 62.



MENSAJE: No se ha podido analizar el sentimiento de los tweet. El 37.0 de los tweets se han definido con polaridad NULA.. NOTA: 7

Figura 62. Análisis del sentimiento de @SergioRG92

### Análisis realizados – actividad

No se encuentran picos de actividad en los tweets del usuario. Esta es una de las características más importante a la hora de clasificar usuarios bot y el usuario no la cumple. Ver figura 63.

MENSAJE: No se han encontrado evidencias de picos de actividad en los tweets del usuario. NOTA: 2

Figura 63. Análisis de la actividad de @SergioRG92

### Análisis realizados – cuenta del usuario

Por último, el análisis del perfil nos aporta datos importantes como la reputación de la cuenta y la proporción de seguidos entre los seguidores. Ambas obtienen buenos resultados en el análisis 64.

MENSAJE: Buena reputación de la cuenta: 0.38%. NOTA: 5  
MENSAJE: El Timeline del usuario no contiene tweets duplicados. NOTA: 3  
MENSAJE: Equilibrada proporción de seguidos. El 0.6285714285714286%. NOTA: 3  
MENSAJE: Proporción baja de RT en los tweets analizados. El 0%. NOTA: 3  
MENSAJE: El usuario no tiene como foto principal la foto por defecto de Twitter. NOTA: 3  
MENSAJE: El usuario posee descripción. NOTA: 3  
MENSAJE: Geolocalización activa. NOTA: 3

Figura 64. Análisis completo de la cuenta @SergioRG92

CUENTA ADMINISTRADA -> NOTA: 4.7

### Conclusiones del análisis

El análisis muestra, lo que parece ser un usuario real que administra su cuenta y distribuye contenido de diferentes fuentes. Existe una diversidad de temas hablados en su perfil y no presenta duplicados de tweets. Tampoco picos de actividad sospechosos en la publicación de contenido. Además, la cuenta tiene una buena reputación y proporción de seguidos.

## **USUARIO HUMANO 2 - @Caiquet11**

Nos encontramos otro usuario humano de la red social Twitter escogido por su alta actividad en Twitter y participación en diversos temas de actualidad. Ha presentado los siguientes resultados en el análisis realizado por TBotDetector.

### **Análisis realizados – set de tweets**

El primero de los análisis indica que el usuario presenta una distribución de tweets libre de duplicados. Primera de las conductas no bot más importante del análisis. Ver figura 65.

MENSAJE: El timeline del usuario no contiene tweets duplicados. NOTA: 3

Figura 65. Análisis de los tweets repetidos @Caiquet11

### **Análisis realizados – proporciones**

Los análisis relacionados con la proporción de las siguientes características que se visualizan en la figura 66 arrojan resultados bajos lo que tiene como consecuencia una baja puntuación como conducta bot. El usuario tiende postear tweets sin utilizar en exceso menciones a otros usuarios, hashtags o enlaces.

MENSAJE: Baja proporción de RT. El 18.18%. NOTA: 2

MENSAJE: Baja proporción de Hashtags. El 9.09%. NOTA: 2

MENSAJE: Baja proporción de URL. El 18.18%. NOTA: 2

MENSAJE: Equilibrada proporción de menciones. El 27.27%. NOTA: 4

Figura 66. Proporciones destacadas de @Caiquet11

### **Análisis realizados – entidades**

En función de los resultados obtenidos en el análisis de proporciones, destacar que no existe ninguna entidad que destaque por el número de apariciones. Las más utilizadas por el usuario son menciones o hashtags y su impacto es leve. Ver figura 67.

MENSAJE: Una de las entidades mas destacadas @queenviolet85, con un total de 11 apariciones, son menciones a otro usuario. NOTA: 0  
MENSAJE: Una de las entidades mas destacadas @Hobby\_Consolas, con un total de 10 apariciones, son menciones a otro usuario. NOTA: 0  
MENSAJE: Una de las entidades mas destacadas, #Ramblas con un total de 8 apariciones, es un Hashtag. NOTA: 0  
MENSAJE: Una de las entidades mas destacadas @policia, con un total de 8 apariciones, son menciones a otro usuario. NOTA: 0  
MENSAJE: No existe una entidad representativa dentro de los tweets analizados. NOTA: 3

Figura 67. Análisis de las entidades principales de @Caiquet11

## Análisis realizados – categorías

La categorización de los tweets ha sido diversa teniendo categorías tan opuestas como son **policía y justicia** o **deporte**. No ha podido categorizar más del 25% de los tweets analizados. Ver figura 68.

```
MENSAJE: Sólo el 9.15% de los tweets se han podido categorizar como policía y justicia. NOTA: 0
MENSAJE: Sólo el 5.88% de los tweets se han podido categorizar como catástrofes y accidentes. NOTA: 0
MENSAJE: Sólo el 3.92% de los tweets se han podido categorizar como ciencia y tecnología. NOTA: 0
MENSAJE: Sólo el 2.61% de los tweets se han podido categorizar como economía y finanzas. NOTA: 0
MENSAJE: Sólo el 1.96% de los tweets se han podido categorizar como deporte. NOTA: 0
MENSAJE: Timeline del usuario con diversidad de temas. NOTA: 3
```

Figura 68. Análisis de las categorías principales de @Caiquet11

## Análisis realizados – sentimiento

Un alto porcentaje de tweets con polaridad nula indica que no ha podido analizar el sentimiento correctamente. Esto puede ser debido al uso de expresiones o mensajes con un leve nivel lingüístico o al uso de expresiones o emoticonos no reconocibles por el análisis 69.

```
MENSAJE: No se ha podido analizar el sentimiento de los tweet. El 76.0 de los tweets se han definido con polaridad NULA.. NOTA: 7
```

Figura 69. Análisis de sentimiento de @Caiquet11

## Análisis realizados – actividad

La actividad del perfil es irregular y retorna un análisis positivo para el análisis de la frecuencia de tweets por día y hora. Ver figura 70.

```
MENSAJE: No se han encontrado evidencias de picos de actividad en los tweets del usuario. NOTA: 2
```

Figura 70. Análisis de la actividad de @Caiquet11

## Análisis realizados – cuenta del usuario

Por último, el análisis del perfil aporta una buena reputación de cuenta y una proporción de seguidos coherente para el perfil. El resultado final de la figura 71 indica que la cuenta no es un usuario bot y se encuentra administrada por un usuario siendo más baja de lo normal con respecto a otros análisis.

```
MENSAJE: Buena reputación de la cuenta: 0.43%. NOTA: 5
MENSAJE: El Timeline del usuario no contiene tweets duplicados. NOTA: 3
MENSAJE: Equilibrada proporción de seguidos. Por cada 1 usuario que sigue, le siguen una media de 0.75. NOTA: 3
MENSAJE: Proporción baja de RT en los tweets analizados. El 66.0%. NOTA: 3
MENSAJE: El usuario no tiene como foto principal la foto por defecto de Twitter. NOTA: 3
MENSAJE: El usuario posee descripción. NOTA: 3
MENSAJE: Geolocalización activa. NOTA: 3
CUENTA ADMINISTRADA -> NOTA: 3.3
```

Figura 71. Análisis general de la cuenta de @Caiquet11

## Conclusiones del análisis

Los informes señalan que la cuenta puede ser una cuenta auténtica de un usuario humano. No contiene tweets duplicados, el uso de entidades es leve y no presenta una actividad en sus publicaciones propia de un usuario bot. También cuenta con buena nota la reputación de la cuenta, siendo equilibrada al igual que la proporción de seguidos obtenida.

### **HIPÓTESIS HUMANO 3 - @TraedRuffles**

Cuenta dedicada al entretenimiento y humor compartida con el resto de usuarios. Ha presentado los siguientes resultados en el análisis realizado por TBotDetector.

#### **Análisis realizados – set de tweets**

Se obtuvo duplicados en el conjunto de tweets del usuario. En la figura 66 se puede observar que estos tweets son dos y su número de apariciones es muy pequeño. Aunque resalta el contenido ya que es una simple respuesta sin ningún propósito adicional.

Sin embargo, el análisis detecta duplicados y le asigna una nota elevada ya que lo interpreta como conducta bot. Figura 72.

```
MENSAJE: El timeline del usuario contiene tweets duplicados. NOTA: 10  
TWEET REPETIDO: - No. | con un total de 3 repeticiones  
TWEET REPETIDO: - Sí. | con un total de 3 repeticiones
```

Figura 72. Lista de tweets repetidos de @TraedRuffles

#### **Análisis realizados – proporciones**

En el segundo análisis se empieza a distinguir la conducta no bot del perfil. Escaso número de menciones a otros usuarios al igual que de RT de la cuenta. La proporción de Hashtags y de enlaces destaca por no contener ni una de estas dos características en los 250 tweets analizados. Ver figura 73.

```
MENSAJE: Baja proporción de RT. El 0.0%. NOTA: 2  
MENSAJE: Baja proporción de Hashtags. El 0.0%. NOTA: 2  
MENSAJE: Baja proporción de URL. El 0.0%. NOTA: 2  
MENSAJE: Baja proporción de menciones. El 9.09%. NOTA: 2
```

Figura 73. Proporciones principales de @TraedRuffles

## Análisis realizados – entidades

Los resultados anteriores confirman el siguiente análisis. Tan sólo una entidad ha podido ser obtenida con un número de repeticiones mayor que uno. Por lo tanto, no existen entidades representativas en el usuario. Ver figura 74.

MENSAJE: Una de las entidades mas destacadas @jon\_mcenroe, con un total de 5 apariciones, son menciones a otro usuario. NOTA: 0  
MENSAJE: No existe una entidad representativa dentro de los tweets analizados. NOTA: 3

Figura 74. Análisis de las principales entidades de @TraedRuffles

## Análisis realizados – categorías

El 16% ha sido posible de los tweets ha sido posible ser categorizado como Deporte. Aunque los resultados de categorización no sean elevados, el perfil tiene una amplia diversidad de temas como por ejemplo los **deportes**, la **política o el ocio, estilo de vida y tiempo libre**. Ver figura 75.

MENSAJE: El 16.66% de los tweets se han podido categorizar como deporte. NOTA: 0  
MENSAJE: Sólo el 2.63% de los tweets se han podido categorizar como política. NOTA: 0  
MENSAJE: Sólo el 1.75% de los tweets se han podido categorizar como arte y cultura. NOTA: 0  
MENSAJE: Sólo el 1.75% de los tweets se han podido categorizar como asuntos sociales. NOTA: 0  
MENSAJE: Sólo el 0.88% de los tweets se han podido categorizar como ocio, estilo de vida y tiempo libre. NOTA: 0  
MENSAJE: Timeline del usuario con diversidad de temas. NOTA: 3

Figura 75. Análisis de las principales categorías de @TraedRuffles

## Análisis realizados – sentimiento

El análisis de sentimiento proporciona una polaridad nula para más del 50% de los mensajes. Esto puede ser al uso de la cuenta a la publicación de frases humorísticas o que carecen de un sentido común o de realidad para el análisis. Ver figura 76.

MENSAJE: No se ha podido analizar el sentimiento de los tweet. El 56.0 de los tweets se han definido con polaridad NULA.. NOTA: 7

Figura 76. Análisis del sentimiento de @TraedRuffles

## **Análisis realizados – actividad**

Existen picos de actividad en la cuenta. Este análisis es coherente sabiendo que la cuenta pública de manera continua todos los días para el entretenimiento del resto de usuarios que son seguidores. Sin embargo, al igual que ocurre en el análisis de tweets duplicados, el algoritmo detecta esta tendencia y la cataloga como bot. Ver figura 77.

MENSAJE: Frecuencia elevada de tweets en el día Miércoles, a la hora 23. NOTA: 0  
MENSAJE: Frecuencia elevada de tweets en el día Jueves, a la hora 21. NOTA: 0  
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Miércoles se escribieron 15 tweets. NOTA: 0  
MENSAJE: Se han encontrado picos de actividad continua en una determinada hora del día. NOTA: 10

**Figura 77. Análisis de la actividad de @TraedRuffles**

## **Análisis realizados – cuenta del usuario**

Por último, el análisis del perfil del usuario. Destaca que el usuario tenga una mala reputación de la cuenta debido al alto número de seguidores respecto a los que él sigue. Al igual que de cada 1 seguido de la cuenta, tiene 281 seguidores. Esto sucede por la popularidad de la cuenta dentro de la red social. Ver figura 78.

MENSAJE: Mala reputación de la cuenta: 1.0%. NOTA: 10  
MENSAJE: El Timeline del usuario contiene tweets duplicados. NOTA: 10  
MENSAJE: Proporción elevada de seguidos. Por cada 1 usuario que sigue, le siguen una media de 281.13. NOTA: 10  
MENSAJE: Proporción baja de RT en los tweets analizados. El 8.0%. NOTA: 3  
MENSAJE: El usuario no tiene como foto principal la foto por defecto de Twitter. NOTA: 3  
MENSAJE: El usuario posee descripción. NOTA: 3  
MENSAJE: Geolocalización no activa. NOTA: 6  
CUENTA ADMINISTRADA -> NOTA: 4.6

**Figura 78. Análisis general de la cuenta de @TraedRuffles**

## **Conclusiones del análisis**

El resultado final que proporciona TBotDetector nos indica que la cuenta podría ser administrada por un usuario y no un bot. Esto quiere decir que se tratan diferentes temas con frecuencia y participativo en la red social, pero el perfil no tiene una intencionalidad más allá de informar y expresar su opinión personal.

Destacar, que dos de los análisis más importantes y determinantes de TBotDetector han arrojado resultados bot. Sin embargo, el estudio de características como el uso de menciones, hashtags, enlaces a otras páginas y la proporción de RT, unido a un potente análisis de entidades y de clasificación de textos han podido declinar el análisis para no obtener resultados erróneos.

## 5.2 ANÁLISIS MÚLTIPLE

Los experimentos para el análisis se han seguido también los requisitos definidos en el anterior capítulo de la presente memoria. Se han tomado comparaciones de un único usuario para optimizar el tiempo de procesamiento. Sin embargo, si se desea se pueden realizar análisis de más usuarios. El algoritmo está preparado para que muestre los resultados por cada usuario.

A diferencia del análisis simple, en esta ocasión y teniendo conocimiento de que el usuario con el que se va a comparar es un usuario bot, el análisis realiza comparaciones de las características del bot con el usuario. Sólo si encuentra semejanzas con todas las características del usuario bot, se puede afirmar que estamos ante otro perfil con propósitos idénticos. Queda para futuras implementaciones, decidir si el usuario puede pertenecer a otra red que no sea la del primer usuario y evaluar la posibilidad de crear un análisis para obtener el porcentaje de confianza obtenido dado los resultados.

### 5.2.1 EXPERIMENTOS CON REDES BOT

#### OBJETIVOS

Detectar redes de bots programadas para un determinado propósito. Estudiar las similitudes entre perfiles en base a sus características para obtener una decisión final que permita etiquetar el usuario en la misma red bot o no.

#### DATOS

Archivo con el usuario bot y el usuario que se desea estudiar y comprobar si pertenece a la misma red de bots. De cada usuario se analizarán características de su contenido con un set de 250 tweets por cada uno.

#### **HIPÓTESIS RED BOT 1: @blogAceiteoliva, @BlogEcologismo**

El primer análisis múltiple corresponde a dos usuarios que pertenecen presuntamente a una misma red de bots. TBotDetector ha obtenido los siguientes resultados.

## Análisis realizados – comparación set de tweets

Coincidencias en numerosos tweets que han sido repetidos por ambos usuarios. Se puede observar el tweet más el número de repeticiones de cada usuario. Es importante resaltar que incluso el número de repeticiones es idéntico. Ver figura 79.

```
MENSAJE: El usuario bot y el usuario analizado comparten el tweet RT @HumorRisas 10 animales que se han colado en fotos que te harán reír https://t.co/FbMbK3ZDPH https://t.co/YYFJ2aZdLf
El usuario bot Aceite de Oliva con 5 repeticiones
El usuario analizado Ecologismo con 4 repeticiones
. CONCLUSION: RED BOTS
MENSAJE: El usuario bot y el usuario analizado comparten el tweet RT @GuapasFamosas Las 10 fotos mas sexys de Scarlett Johansson https://t.co/oRuq2xTGRc https://t.co/scFzj2HxJ4
El usuario bot Aceite de Oliva con 4 repeticiones
El usuario analizado Ecologismo con 4 repeticiones
. CONCLUSION: RED BOTS
MENSAJE: El usuario bot y el usuario analizado comparten el tweet RT @OcioUltimate Los 15 grandes descubrimientos para ir a Mercadona https://t.co/1NOg7y6txu https://t.co/00cRKvL5iC
El usuario bot Aceite de Oliva con 4 repeticiones
El usuario analizado Ecologismo con 4 repeticiones
. CONCLUSION: RED BOTS
MENSAJE: El usuario bot y el usuario analizado comparten el tweet RT @Sexologianet Los 10 récords sexuales más locos del mundo https://t.co/VLKENQQE4g https://t.co/nfocuhLUVC
El usuario bot Aceite de Oliva con 3 repeticiones
El usuario analizado Ecologismo con 3 repeticiones
. CONCLUSION: RED BOTS
MENSAJE: El usuario bot y el usuario analizado comparten el tweet RT @TeleBLOG 15 fotos polémicas de Sofía (GH16) https://t.co/qyXw5hre2g https://t.co/0XiB2HCWGQ
El usuario bot Aceite de Oliva con 3 repeticiones
El usuario analizado Ecologismo con 4 repeticiones
. CONCLUSION: RED BOTS
MENSAJE: El usuario bot y el usuario analizado comparten el tweet RT @Sexologianet ¿Qué es el pegging? Conoce esta practica poco común https://t.co/16tkv5nVtt https://t.co/06hVxNwMUT
El usuario bot Aceite de Oliva con 3 repeticiones
El usuario analizado Ecologismo con 3 repeticiones
. CONCLUSION: RED BOTS
MENSAJE: El usuario bot y el usuario analizado comparten el tweet RT @GuapasFamosas 10 imágenes que demuestran que Scarlett Johansson gana con los años https://t.co/KPsXEeIzzz https://t.co/6G04jE3jqf
El usuario bot Aceite de Oliva con 3 repeticiones
El usuario analizado Ecologismo con 3 repeticiones
. CONCLUSION: RED BOTS
MENSAJE: El usuario bot y el usuario analizado comparten el tweet RT @OcioUltimate Los desplantas públicos de Letizia https://t.co/Pt3grBQ0zQ https://t.co/0kzthZ01Jo
El usuario bot Aceite de Oliva con 3 repeticiones
El usuario analizado Ecologismo con 3 repeticiones
. CONCLUSION: RED BOTS
MENSAJE: El usuario bot y el usuario analizado comparten el tweet RT @BipolarSuelto 10 celebrities que tienen una larga lista de ex https://t.co/A4YsKxJrLK https://t.co/bbXCAFP1BU
El usuario bot Aceite de Oliva con 2 repeticiones
El usuario analizado Ecologismo con 2 repeticiones
. CONCLUSION: RED BOTS
MENSAJE: El usuario bot y el usuario analizado comparten el tweet RT @OcioUltimate Mentiras que todos decimos en la intimidad https://t.co/0BokuNLTbB https://t.co/Eh11TzxHQI
El usuario bot Aceite de Oliva con 2 repeticiones
El usuario analizado Ecologismo con 2 repeticiones
. CONCLUSION: RED BOTS
```

Figura 79. Lista de tweets semejantes entre los usuarios analizados

## Análisis realizados – comparación proporciones

El segundo análisis muestra los porcentajes de diferentes características y la diferencia que existe en comparación con el usuario ya analizado y que se comportaba como un bot. De nuevo, los resultados son calcados entre ambas cuentas. Ver figura 80.

```
MENSAJE: El usuario analizado tiene un 100.0% de RT en su timeline.
La diferencia con el usuario bot es apenas del 0.0%. CONCLUSION: RED BOTS
MENSAJE: El usuario analizado tiene un 100.0% de uso de URL en su timeline.
La diferencia con el usuario bot es apenas del 2.5%. CONCLUSION: RED BOTS
MENSAJE: El usuario analizado tiene un 0.0% de uso de HASHTAGS en su timeline.
La diferencia con el usuario bot es apenas del 0.0%. CONCLUSION: RED BOTS
```

Figura 80. Comparación de proporciones entre los usuarios del análisis



## **Análisis realizados – comparación entidades**

El análisis de entidades, al igual que ocurría con los tweets, aporta valores idénticos en las entidades. De nuevo, otro análisis que indica un comportamiento idéntico. Ver figura 81.

```
MENSAJE: La entidad @GuapasFamosas coincide con una de las entidades del usuario bot.  
Número total de apariciones en tweets de la entidad en el usuario analizado: 78  
Número total de apariciones en tweets de la entidad en el usuario bot: 97. CONCLUSION: RED BOTS  
MENSAJE: La entidad @OcioUltimate coincide con una de las entidades del usuario bot.  
Número total de apariciones en tweets de la entidad en el usuario analizado: 29  
Número total de apariciones en tweets de la entidad en el usuario bot: 36. CONCLUSION: RED BOTS  
MENSAJE: La entidad @TeleBLOG coincide con una de las entidades del usuario bot.  
Número total de apariciones en tweets de la entidad en el usuario analizado: 31  
Número total de apariciones en tweets de la entidad en el usuario bot: 31. CONCLUSION: RED BOTS  
MENSAJE: La entidad @BipolarSuelto coincide con una de las entidades del usuario bot.  
Número total de apariciones en tweets de la entidad en el usuario analizado: 25  
Número total de apariciones en tweets de la entidad en el usuario bot: 25. CONCLUSION: RED BOTS  
MENSAJE: La entidad Scarlett Johansson coincide con una de las entidades del usuario bot.  
Número total de apariciones en tweets de la entidad en el usuario analizado: 17  
Número total de apariciones en tweets de la entidad en el usuario bot: 17. CONCLUSION: RED BOTS  
MENSAJE: La entidad @Sexologianet coincide con una de las entidades del usuario bot.  
Número total de apariciones en tweets de la entidad en el usuario analizado: 13  
Número total de apariciones en tweets de la entidad en el usuario bot: 13. CONCLUSION: RED BOTS
```

Figura 81. Comparación de entidades entre los usuarios aportando gran valor al análisis

## **Análisis realizados – comparación categorías**

Por último, se han detectado las mismas categorías extraídas por el análisis de textos. Tanto el usuario bot como el usuario analizado obtienen los mismos resultados y estas son las cinco categorías más importantes. Ver figura 82.

```
MENSAJE: Clasificación de los tweets ocio, estilo de vida y tiempo libre coincide con la clasificación dada al usuario bot  
Número total de tweets clasificados en el usuario analizado: 4.0  
Número total de tweets clasificados en el usuario bot: 4.0. CONCLUSION: RED BOTS  
MENSAJE: Clasificación de los tweets asuntos sociales coincide con la clasificación dada al usuario bot  
Número total de tweets clasificados en el usuario analizado: 4.0  
Número total de tweets clasificados en el usuario bot: 4.0. CONCLUSION: RED BOTS  
MENSAJE: Clasificación de los tweets salud coincide con la clasificación dada al usuario bot  
Número total de tweets clasificados en el usuario analizado: 3.0  
Número total de tweets clasificados en el usuario bot: 4.0. CONCLUSION: RED BOTS  
MENSAJE: Clasificación de los tweets religión y credos coincide con la clasificación dada al usuario bot  
Número total de tweets clasificados en el usuario analizado: 3.0  
Número total de tweets clasificados en el usuario bot: 3.0. CONCLUSION: RED BOTS  
MENSAJE: Clasificación de los tweets deporte coincide con la clasificación dada al usuario bot  
Número total de tweets clasificados en el usuario analizado: 1.0  
Número total de tweets clasificados en el usuario bot: 1.0. CONCLUSION: RED BOTS
```

Figura 82. Comparación de categorías que aportan gran valor al análisis entre usuarios al obtener valores idénticos

## **Análisis realizados – comparación actividad**

El análisis de la actividad de los usuarios de nuevo proporciona resultados idénticos. Esto indica que los bot actúan conjuntamente en unas determinadas horas y días concretos. Ver figura 83.

MENSAJE: Existen picos de actividad en uno o varios días y a una determinada hora entre el usuario analizado y el usuario bot.

```
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Domingo se escribieron 19 tweets. CONCLUSION: -----
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Domingo se escribieron 17 tweets. CONCLUSION: -----
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Domingo se escribieron 18 tweets. CONCLUSION: -----
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Domingo se escribieron 26 tweets. CONCLUSION: -----
MENSAJE: Se han encontrado picos de actividad a una determinada hora de un mismo día. CONCLUSION: -----
-----
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Domingo se escribieron 19 tweets. CONCLUSION: -----
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Domingo se escribieron 17 tweets. CONCLUSION: -----
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Domingo se escribieron 18 tweets. CONCLUSION: -----
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Domingo se escribieron 26 tweets. CONCLUSION: -----
MENSAJE: Se han encontrado picos de actividad a una determinada hora de un mismo día. CONCLUSION: -----
PERTENECEN A LA MISMA RED DE BOTS
```

**Figura 83. Comparación de actividad entre los usuarios que arroja resultados exactos entre los usuarios**

## **Conclusión del análisis**

Los dos usuarios guardan coincidencias en todas las categorías analizadas. Los resultados arrojan similitudes en determinados casos del 100%. Resultado que señalaría probable la compartición de un determinado propósito dentro de la red social, llegando a estar automatizados conjuntamente para determinadas horas y días duplicando los mismos tweets. El usuario @BlogEcologismo pertenece presuntamente a la misma red que @blogAceiteOliva.

## **RED BOT 2: @moralesr904, @LuisDavidPz**

### **Análisis realizados – comparación set de tweets**

El análisis no detectó tweets repetidos entre ambos usuarios. Los clústeres analizados tenían coincidencias entre sí ya que ha sido posible realizar un análisis del usuario, pero no se han detectado tweets duplicados semejantes entre ambos perfiles

### **Análisis realizados – comparación de proporciones**

El segundo análisis utiliza dos usuarios sospechosos de pertenecer a una red de bots que imponen tendencias en Twitter sobre política.

Se trata de una red más sofisticada que la anterior, ya que los tweets contienen son de mayor variedad y sin duplicados. Sin embargo, la tendencia de

ambos a destacar unas determinadas entidades por encima de otras unido a otros análisis, hacen que TBotDetector consiga identificar que todos los perfiles pertenecen probablemente a la misma red.

En el primer análisis destaca la diferencia de RT entre ambas cuentas, apenas un 7,5%. Sin embargo, se obtienen resultados menos precisos que en el anterior análisis debido a la variedad del conjunto de tweets analizados en ambos usuarios. Sin embargo, el análisis no muestra en esta imagen la diferencia del uso de Hashtags entre cuentas. Este resultado ha sido del 15.7%, una cifra que no es elevada pero no entra dentro de las medidas para filtrar del algoritmo. Aunque no sea detectada en esta imagen, si será visible en la vista y no desvirtúa el análisis. Ver figura 84.

```
MENSAJE: El usuario analizado tiene un 42.857142857142854% de RT en su timeline.  
La diferencia con el usuario bot es apenas del 12.86%.. CONCLUSION: RED BOTS  
MENSAJE: El usuario analizado tiene un 0.0% de uso de URL en su timeline.  
La diferencia con el usuario bot es apenas del 10.0%.. CONCLUSION: RED BOTS  
MENSAJE: El usuario analizado tiene un 57.142857142857146% de uso de MENCIONES en su timeline.  
La diferencia con el usuario bot es apenas del 2.86%.. CONCLUSION: RED BOTS
```

Figura 84. Comparación de porcentajes entre los usuarios analizados

## Análisis realizados – comparación de entidades

Para esta red, el análisis de entidades resulta imprescindible ya que anteriormente no se han encontrado tweets repetidos entre los perfiles y los porcentajes obtenidos son menos precisos que en otros análisis. Se aprecia cómo ambos usuarios emplean entidades y Hashtags idénticos con un alto número de apariciones en los respectivos perfiles.

Las entidades mostradas están involucradas con el propósito principal de esta presunta red de bots, destacando la entidad **Venezuela** y el hashtag **#LaidyGómezALa1** y su nombre de perfil **@laidygozmezf**. Ver la figura 85 para este análisis.

```
MENSAJE: La entidad @laidygozmezf coincide con una de las entidades del usuario bot.  
Número total de apariciones en tweets de la entidad en el usuario analizado: 38  
Número total de apariciones en tweets de la entidad en el usuario bot: 65. CONCLUSION: RED BOTS  
MENSAJE: La entidad #LaidyGómezALa1 coincide con una de las entidades del usuario bot.  
Número total de apariciones en tweets de la entidad en el usuario analizado: 15  
Número total de apariciones en tweets de la entidad en el usuario bot: 40. CONCLUSION: RED BOTS  
MENSAJE: La entidad Táchira coincide con una de las entidades del usuario bot.  
Número total de apariciones en tweets de la entidad en el usuario analizado: 15  
Número total de apariciones en tweets de la entidad en el usuario bot: 34. CONCLUSION: RED BOTS  
MENSAJE: La entidad @DarvinsonRojas coincide con una de las entidades del usuario bot.  
Número total de apariciones en tweets de la entidad en el usuario analizado: 31  
Número total de apariciones en tweets de la entidad en el usuario bot: 28. CONCLUSION: RED BOTS  
MENSAJE: La entidad Venezuela coincide con una de las entidades del usuario bot.  
Número total de apariciones en tweets de la entidad en el usuario analizado: 16  
Número total de apariciones en tweets de la entidad en el usuario bot: 16. CONCLUSION: RED BOTS  
MENSAJE: La entidad @diana_dagostino coincide con una de las entidades del usuario bot.  
Número total de apariciones en tweets de la entidad en el usuario analizado: 15  
Número total de apariciones en tweets de la entidad en el usuario bot: 15. CONCLUSION: RED BOTS
```

Figura 85. Comparación de entidades entre los usuarios

## Análisis realizados – comparación de categorías

Al igual que en el anterior análisis, la clasificación de textos representados en la figura 86 nos ayuda a mejorar el análisis entre las cuentas. Todas las categorías en las que se han dividido los tweets permiten identificar la intención de la presunta red o guardan una estrecha relación entre sí. Los tweets han sido categorizados en temas como **política, asuntos sociales, disturbios, conflictos y guerra o policía y justicia**. Ver figura 86.

```
MENSAJE: Clasificación de los tweets política coincide con la clasificación dada al usuario bot
Número total de tweets clasificados en el usuario analizado: 21.0
Número total de tweets clasificados en el usuario bot: 51.0. CONCLUSION: RED BOTS
MENSAJE: Clasificación de los tweets policía y justicia coincide con la clasificación dada al usuario bot
Número total de tweets clasificados en el usuario analizado: 22.0
Número total de tweets clasificados en el usuario bot: 25.0. CONCLUSION: RED BOTS
MENSAJE: Clasificación de los tweets trabajo coincide con la clasificación dada al usuario bot
Número total de tweets clasificados en el usuario analizado: 12.0
Número total de tweets clasificados en el usuario bot: 19.0. CONCLUSION: RED BOTS
MENSAJE: Clasificación de los tweets asuntos sociales coincide con la clasificación dada al usuario bot
Número total de tweets clasificados en el usuario analizado: 13.0
Número total de tweets clasificados en el usuario bot: 15.0. CONCLUSION: RED BOTS
MENSAJE: Clasificación de los tweets disturbios, conflictos y guerra coincide con la clasificación dada al usuario bot
Número total de tweets clasificados en el usuario analizado: 7.0
Número total de tweets clasificados en el usuario bot: 11.0. CONCLUSION: RED BOTS
```

Figura 86. Comparación de categorías entre los usuarios

## Análisis realizados – comparación de la actividad

Por último, la actividad de ambas cuentas es común. Existen pocas variaciones entre los tweets publicados cada día y se han encontrado picos de actividad a determinadas horas en ambas cuentas. Por lo que las cuentas están organizadas para que en determinadas horas y días aumenten la actividad y tengan un mayor posicionamiento en la red. Ver figura 87.

```
MENSAJE: Existen picos de actividad en uno o varios días y a una determinada hora entre el usuario analizado y el usuario bot. CONCLUSION: RED BOTS
```

```
MENSAJE: Frecuencia elevada de tweets en el día Viernes, a la hora 5. CONCLUSION: -----
MENSAJE: Frecuencia elevada de tweets en el día Jueves, a la hora 4. CONCLUSION: -----
MENSAJE: Frecuencia elevada de tweets en el día Viernes, a la hora 19. CONCLUSION: -----
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Lunes se escribieron 25 tweets. CONCLUSION: -----
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Martes se escribieron 29 tweets. CONCLUSION: -----
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Miércoles se escribieron 14 tweets. CONCLUSION: -----
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Jueves se escribieron 26 tweets. CONCLUSION: -----
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Viernes se escribieron 16 tweets. CONCLUSION: -----
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Sábado se escribieron 19 tweets. CONCLUSION: -----
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Domingo se escribieron 21 tweets. CONCLUSION: -----
-----
MENSAJE: Frecuencia elevada de tweets en el día Viernes, a la hora 5. CONCLUSION: -----
MENSAJE: Frecuencia elevada de tweets en el día Viernes, a la hora 19. CONCLUSION: -----
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Lunes se escribieron 22 tweets. CONCLUSION: -----
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Martes se escribieron 24 tweets. CONCLUSION: -----
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Miércoles se escribieron 13 tweets. CONCLUSION: -----
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Jueves se escribieron 25 tweets. CONCLUSION: -----
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Viernes se escribieron 28 tweets. CONCLUSION: -----
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Sábado se escribieron 21 tweets. CONCLUSION: -----
MENSAJE: Frecuencia elevada de tweets en un mismo día. El Domingo se escribieron 18 tweets. CONCLUSION: -----
MENSAJE: Se han encontrado picos de actividad a una determinada hora de un mismo día. CONCLUSION: -----
PERTENECEN A LA MISMA RED DE BOTS
```

Figura 87. Comparación de actividad entre usuarios que refleja una similitud en sus publicaciones

## Conclusión del análisis

Finalmente, como se aprecia en la figura 87, el análisis ha podido dar un resultado positivo y la cuenta @LuisDavidPz podría ser bot y pertenecer a una misma red. Estas cuentas parecen estar organizadas para desacreditar. Es normal que ambas guarden similitud en su actividad, entidades y categorías ya que impulsan una determinada postura.

La cuenta pretende esconder su realidad identidad cuidando los tweets publicados para no duplicar el contenido y siempre aportar valor en cada publicación.

### 5.2.2 EXPERIMENTOS SIN REDES BOT

Cuando un usuario no presenta semejanzas en la comparación de clústeres con el usuario bot el algoritmo no sigue evaluando el resto de análisis.

En los diagramas de actividad se puede apreciar como si los clústeres no coinciden, el algoritmo no sigue obteniendo más datos. Sin embargo, los análisis del usuario bot si se obtienen.

Esto es debido a la potencia que aporta la extracción de clústeres de cada usuario. Dada su potente agrupación de tweets semejantes dada una característica en común, permite discriminar del análisis aquellos usuarios que no posean clústeres con títulos semejantes al del usuario bot.

La figura 88 muestra el mensaje mostrado cuando un usuario no pertenece a la misma red de bots. El análisis es independiente para cada usuario que se considere analizar. Esto permite obtener usuarios que, si puedan pertenecer a la misma red de bots, cómo también usuarios que no pueden ser considerados pertenecientes a la misma red.

`MENSAJE: No existen coincidencias entre el contenido de ambos perfiles. CONCLUSION: NO RED BOTS`

Figura 88. Mensaje si los usuarios no pertenecen a la misma red bot

## OBJETIVOS

Obtener análisis negativos en la obtención de redes bot dado un conjunto de usuarios que no pertenecen a la misma red bot o usuarios que reales de la aplicación.

## DATOS

Se han analizado las características de los usuarios aportados para cada análisis. Cada uno de ellos con una muestra de 250 tweets.

**HIPÓTESIS DE USUARIOS NO PERTENECIENTES A LA MISMA RED1:**  
**@afrasecool y @SergioRG92**

El primer caso, se analizaron los usuarios **@afrasecool** y **@SergioRG92**. No compartían ni la característica de bot ya que el usuario analizado es un usuario humano cómo se ha podido comprobar en el análisis simple detallado en esta memoria.

#### **HIPÓTESIS DE USUARIOS NO PERTENECIENTES A LA MISMA RED 2: @ManuelAntonioFD y @SoniaVazquez75**

El siguiente análisis fue realizado con dos cuentas aparentemente pertenecientes a la misma red de bots, pero los resultados obtenidos no fueron los deseados. Estos presuntos usuarios son **@ManuelAntonioFD** y **@SoniaVazquez75**, ambos pertenecientes a una presunta red de bots. El primero de ellos mantiene una actividad constante en la publicación de tweets, mientras que el usuario que se desea conocer su implicación en la red lleva sin actividad desde el 26 de junio del 2017.

El algoritmo obtiene los últimos 250 tweets de ambos usuarios. Esto tiene la consecuencia de que si un usuario no es activo el análisis no va a arrojar resultados satisfactorios. Sin embargo, se puede dar la posibilidad de que la cuenta haya abandonado su intención dentro de la red de bots y este desuso implique la no consideración como cuenta perteneciente a la misma red.

Por lo tanto, para este análisis, el abandono de la cuenta por un prolongado periodo de tiempo supone que no es apta para el análisis y por lo tanto no es considerada como cuenta perteneciente a la red de bots.

#### **HIPÓTESIS DE USUARIOS NO PERTENECIENTES A LA MISMA RED 2: @afrasecool y @ManuelAntonioFD**

El último ejemplo consta de dos usuarios bot, **@afrasecool** y **@ManuelAntonioFD**, que como se puede apreciar, han sido analizados en los anteriores ejemplos. Estos usuarios si bien son analizados por el análisis simple que propone este proyecto, obtienen resultados bot. Sin embargo, el contenido y propósitos de las cuentas es totalmente diferente.

# Capítulo 6

## 6.1 CONCLUSIÓN

En este trabajo se han presentado métodos para ayudar en la aportación de indicios sobre si un usuario de Twitter es un bot y si dos usuarios bot comparten la misma red. Para ello se han empleado una serie de servicios de análisis del lenguaje natural, el servicio MeaningCloud.

Se ha podido obtener perfiles bot mediante las características descritas e implementadas, sin embargo, es posible aumentar el número de características para que la precisión sea mayor.

También, el análisis para detectar redes bot puede ser ampliado con nuevas comparaciones entre los usuarios. Quedan abiertas otras vías de estudio como el análisis sintáctico de los tweets. Es necesario considerarlo, pues los análisis enfocados a características pueden resultar poco eficientes si la calidad de los bots aumenta.

Estas nuevas características deben implementarse tras un estudio cuidado de las técnicas usadas por el bot. Si bien, cada vez es más difícil detectar estos comportamientos por la escasa diferencia con un usuario humano. Podría llegar un momento en que un bot no pueda distinguirse de un humano sin un gran esfuerzo.

## 6.2 LINEAS FUTURAS

Se plantean a continuación una serie de objetivos para próximas versiones a tener en cuenta.

- Implementar la interfaz gráfica para visualizar los análisis realizados. Es una de las primeras mejoras, permite obtener un mayor potencial de los datos.
- Mejorar los tiempos de análisis. Esta característica es posible mejorarla sustancialmente con la obtención de un plan de pago MeaningCloud para mejorar el número de peticiones por segundo al API.
- Aumentar el número de características para dotar de mayor precisión al análisis.
- Aumentar el grado de precisión en el análisis de redes bot. Diferenciar si los perfiles analizados pueden ser usuarios de otras redes e incluir un sistema de



puntuación para obtener la confianza del análisis similar al análisis simple. La estructura está construida, ya que se utiliza la misma que en el análisis simple.

- Incorporar modelos de clasificación personalizados para TBotDetector. Permitiría optimizar los resultados obtenidos de sus tecnologías haciéndolo más preciso para los análisis.

### 6.3 CONCLUSION

This work introduced methods to gather evidences supporting if a user behaves as a bot, and if two bots shared the same network. For this goal, natural language processing capabilities from MeaningCloud have been used.

It was possible to analyse bots by using the analysis reports as presented in this work. However, it is possible to increase them and achieve better performance.

Also, features can be improved to compare profiles. For example, syntactic analysis of twits. However, feature-driven scans may be ineffective if bot quality increases.

New features should be implemented as bots evolve in Twitter. The analysis to detect bot nets can be expanded with new user comparisons. Current feature based analysis can prove inefficient if the quality of the bots increases. There are other approaches to study such as the syntactic analysis of the tweets that could improve efficacy.



# Bibliografía

- [1] Kit Smith. *44 estadísticas de Twitter para 2016*. (2016) <https://www.brandwatch.com/es/2016/06/44-estadisticas-twitter-2016/>
- [2] Pepo Jiménez, *Así funciona una red de bots en Twitter diseñada para promocionar un medio*. (2016). Recuperado <http://www.vozpopuli.com/memesis/Red de Bots-Twitter-Spam-Red de bots-Que-es 0 881911833.html>
- [3] Farhad Manjoo, *La falsa realidad creada por los bots en Twitter*. (2017) <https://www.nytimes.com/es/2017/06/05/la-falsa-realidad-creada-por-los-bots-en-twitter/>
- [4] *Los jóvenes influyentes del PP gallego*. (2016) <https://botsdetwitter.wordpress.com/category/perfiles-falsos/> bots pp Galicia
- [5] *Una red de bots del Partido Popular hace spam político en Twitter desde Cataluña*. (2016) <https://botsdetwitter.wordpress.com/2015/02/04/una-red-de-bots-del-partido-popular-hace-spam-politico-en-twitter-desde-cataluna/>
- [6] Redacción Tecnosfera, *El 51,8% del tráfico en internet es generado por bots*. (2017) <http://www.eltiempo.com/tecnosfera/novedades-tecnologia/el-51-8-del-trafico-en-internet-es-generado-por-bots-70720>
- [7] Clara Castellón, *20 sorprendentes estadísticas sobre las redes sociales*. (2017) <https://www.fhios.es/20-sorprendentes-estadisticas-sobre-las-redes-sociales/>
- [8] César Salza, *15% de las cuentas de Twitter son bots: estudio*. (2017) <https://www.cnet.com/es/noticias/15-por-ciento-de-las-cuentas-de-twitter-son-bots/>
- [9] Juanjo Santana, *¿Qué son los bots? Tipos y uso*. (2017) <https://www.enredia.es/que-son-los-bots-tipos-usos/> tipos de bots
- [10] @polbots ComProp Research. (2014) [https://twitter.com/polbots?ref\\_src=twsrc%5Etfw&ref\\_url=http%3A%2F%2Fcomprom.oii.ox.ac.uk%2F2017%2F07%2F18%2Fgovernment-cyber-troops-manipulate-facebook-twitter-study-says%2F](https://twitter.com/polbots?ref_src=twsrc%5Etfw&ref_url=http%3A%2F%2Fcomprom.oii.ox.ac.uk%2F2017%2F07%2F18%2Fgovernment-cyber-troops-manipulate-facebook-twitter-study-says%2F)
- [11] *¿Quiénes son los fundadores de Twitter?* (2017) <http://www.masadelante.com/faqs/twitter>
- [12] Onur Varol, Emilio Ferrara, Clayton A. Davis, Filippo Menczer, Alessandro Flammini. *Online Human-Bot Interactions: Detection, Estimation, and Characterization*. Center of Complex Networks and Systems Research, Indiana University, Bloomington, Us InformationSciences Institute, University California, Marina del Rey, CA, US.
- [13] Web oficial de Botometer <https://botometer.iuni.iu.edu/#/>

- [14] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, Alessandro Flammini. *The Rise of Social Bots*. Indiana University
- [15] Genbeta. *El 15% de los usuarios Twitter son bots según un estudio, y les gusta interactuar entre ellos*. (2017) <https://www.genbeta.com/redes-sociales-y-comunidades/el-15-de-los-usuarios-de-twitter-son-bots-segun-un-estudio-y-les-gusta-interactuar-entre-ellos>
- [16] @StatusPeople. *App Fakers*. (2017) <https://fakers.statuspeople.com/>
- [17] @davc and @grossnasty. (2017) <https://www.twitteraudit.com/>
- [18] Luis Sánchez Fernández, Jesus Arias Fisteus, Pablo Basanta-Val, Mariluz Congosto. (2015) *Primero resultados hacia la detección automática de bots en Twitter*. Universidad Carlos III. Departamento de Ingeniería Telemática.
- [19] Josemy Duarte, Gabriel Rodríguez (2016) *Herramienta para el análisis de la interacción, identificación de patrones y clasificación de usuarios en humanos, ciborgs y bot de la red de microblogging Twitter*. Universidad Central de Venezuela, Facultad de Ciencias. Escuela de Computación.
- [20] *Redes de bots en el PSOE de Andalucía para beneficiar a Susana Díaz*. (2017) <https://botsdetwitter.wordpress.com/2016/07/11/psoeandaluz/>
- [21] *blog de análisis de redes bot en la red Social Twitter*. (2017) <https://botsdetwitter.wordpress.com>
- [22] *Arquitectura de las aplicaciones web* (2015) Instituto tecnológico de Matehuala. <https://programacionwebisc.wordpress.com/2-1-arquitectura-de-las-aplicaciones-web/>
- [23] *Web Material Design*. (2017) Google. <https://getmdl.io/>
- [24] *Software de análisis de textos*. (2017) New York. <https://www.meaningcloud.com/es/>
- [25] <https://www.meaningcloud.com/es/soluciones/analisis-de-medios>
- [26] <https://www.meaningcloud.com/es/soluciones/analitica-de-texto> análisis de textos en MeaningCloud
- [27] *Blog Analítica de Medios Sociales* (2017) <https://www.meaningcloud.com/es/productos/extraccion-topics>
- [28] *Modelos de clasificación para MeaningCloud* (2017) <https://www.meaningcloud.com/developer/text-classification/doc/1.1/what-is-model>
- [29] *IBM Watson Data Platform*. IBM Analytics. (2017) [https://www.ibm.com/analytics/es/es/watson-data-platform/platform.html?S\\_PKG=AW&cm\\_mmc=Search\\_Google- -Leadership+Agenda\\_DIGITAL+INBOUND- -ES- -+ibm++watson\\_Broad\\_AW&cm\\_mmca1=000016QS&cm\\_mmca2=10004264&mkwid=d61db356-5e6d-457d-aa1f-80320b624e79|594|10758](https://www.ibm.com/analytics/es/es/watson-data-platform/platform.html?S_PKG=AW&cm_mmc=Search_Google- -Leadership+Agenda_DIGITAL+INBOUND- -ES- -+ibm++watson_Broad_AW&cm_mmca1=000016QS&cm_mmca2=10004264&mkwid=d61db356-5e6d-457d-aa1f-80320b624e79|594|10758) IBM Watson

- [30] Rollins, J. *Why we need a methodology for data science*. (2015) <http://www.ibmbigdatahub.com/blog/why-we-need-methodology-data-science>
- [31] Maimon, O. y Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook, Second Edition*. Springer.
- [32] Web oficial de Thymeleaf. (2017) Thymeleaf Team <http://www.thymeleaf.org/>
- [33] *Integrando Thymeleaf en Spring MVC*. (2017) Blog Tutor de Programación <http://acodigo.blogspot.com.es/2017/04/spring-mvc-integrando-thymeleaf.html>
- [34] ¿Qué son los bots en Twitter? (2016) <http://blog.quantika14.com/blog/2016/06/25/que-son-los-bots-en-twitter/>
- [35] Herramienta detector de bots. (2012) <http://thebotbusters.org/>
- [36] Carlos Pesquera. ¿Qué es un Bean de Spring? (2014) <http://carlospesquera.com/que-es-un-pojo-ejb-y-un-bean/>
- [37] IBM Watson características (2017) <http://www-03.ibm.com/press/es/es/presskit/45119.wss>